УДК 004.56+003.26

МЕТОД СТЕГАНОГРАФИЧЕСКОГО ПРЕОБРАЗОВАНИЯ НА ОСНОВЕ ИЗМЕНЕНИЯ ПРОСТРАНСТВЕННОЙ ОБЛАСТИ РАСТРИРОВАННОГО ДОКУМЕНТА-КОНТЕЙНЕРА

М. Г. Савельева

Белорусский государственный технологический университет,
ул. Свердлова 13а, 220006, г. Минск, Беларусь, saveleva@belstu.by

Представлены метод и алгоритмы стеганографического преобразования, использующие элементы web-приложения на основе растровой графики в качестве контейнера. Основной элемент-контейнер представляет собой пиксель изображения, чьи цветовые параметры модифицируются в цветовой модели при встраивании информации. Процессы внедрения и извлечения данных происходят в пикселях, выбранных на основе полутоновых значений, полученных путем растрирования векторных текстовых символов. Количество цветовых каналов используемых для выбора пикселей и внедрения сообщения, зависит от цветовых характеристик изображения и общего размера передаваемого сообщения. Это позволит учесть особенности изображения и эффективно встраивать информацию в графический контейнер, минимизируя визуальные изменения и сохраняя незначительное влияние на качество отображения.

Ключевые слова: стеганографические методы; растрирование; алгоритм; цвет; пространственная область; авторское право.

В современной эпохе цифровая трансформация значительно изменяет обработку и использование информации, так как все больше данных переходит в электронный формат. Это обстоятельство делает крайне важным хранение, передачу и использование данных в нашей повседневной жизни. Вместе с этим возрастает угроза цифрового пиратства, которое представляет собой несанкционированное копирование и распространение информации. Такая деятельность может привести к серьезным экономическим и правовым последствиям, а также создать проблемы для обеспечения безопасности личных и коммерческих данных [1].

Электронные текстовые документы, подвергающиеся намеренным или ненамеренным изменениям, могут легко терять свое первоначальное состояние, становясь частью растровой или векторной графики. Одной из ключевых проблем текстовых векторных документов является растрирование. При передаче через различные каналы, векторные изображения могут быть преобразованы в растровые без согласования с авторами. Обратный переход в векторную графику возможен только через трассировку или ручную перерисовку всех элементов изображения. Это важно учитывать при разработке математических моделей для стеганографических систем, поскольку они должны быть способны адаптироваться к таким изменениям и сохранять свою эффективность независимо от изменений в исходном текстовом документе-контейнере [2].

Процесс растрирования с заданным разрешением представляет собой создание растровой сетки с фиксированными ячейками, где каждая ячейка представляет пиксель изображения. Затем в этой сетке происходит заполнение (закрашивание) ячеек, в которых находятся точки исходной фигуры изображения. В зависимости от количества точек, попавших в 1 ячейку, она получает свой цвет (в черно-белых изображениях это градиент оттенков от белого к черному) [3]. Кроме того, общее количество пикселей для отображения буквы увеличивается, так как пиксельная сетка должна вместить новые оттенки, что приводит к появлению пиксельных элементов для отображения этих новых оттенков. Эти изменения существенно влияют на визуальное представление исходной информации и могут привести к потере четкости и качества отображения текстовых символов.

В виду того, что невозможно сохранить цвет и размер векторных символов электронного текстового документа в исходном варианте. При конвертации в растровый формат контур буквы начинает «расплываться», цвет по контуру переходит в градиент, при этом общее количество пикселей для отображения буквы увеличивается. Именно в эти новые пиксели можно внедрять информацию *M*. Один из методов, использующих внедрение в полутоновые оттенки, был описан в [4]. Особенностью данного метода является осуществление процессов внедрения / извлечения *M* при сравнительном анализе значений одного или двух цветовых координат базового пикселя и пикселя для внедрения.

Формально процесс встраивания (осаждения) тайных сообщений *M*, с помощью которого, в частности, можно решать задачу защиты авторского права на контент, содержащийся в документах из множества *C*, можно описать как стеганографическую модель [5].

Модель строится на основе следующих положениях. Произвольное тайное сообщение *M* можно скрыть в контейнере *C* при использовании ключей *K*, где $M \in \boldsymbol{M}$, $C \in \boldsymbol{C}$; $K \in \boldsymbol{K}$. Результатом такого преобразования будет стегоконтейнер *S*, $S \in \boldsymbol{S}$ $\{(M_1, C_1, K_1), (M_2, C_2, K_2), \ldots, (M_z, C_z, K_z)\}$, $\boldsymbol{S} \in \{ S_1, S_2, \ldots, S_t \}$. Полагаем, что $\boldsymbol{M}$ − множество скрываемых сообщений, $\boldsymbol{M} = \{M_1, M_2, \ldots, M_n\}$, $\boldsymbol{C}$ − множество контейнеров (в нашем случае − изображения), $\boldsymbol{C} = \{C_1, C_2, \ldots, C_r\}$ $(r > n)$, $\boldsymbol{K}$ – множество всех ключей, $\boldsymbol{K} = \{ K_1, K_2, \ldots, K_a\}$ [6].

Процесс встраивания (осаждения) тайных сообщений $\boldsymbol{M}$, с помощью которого, в частности, можно решать задачу защиты авторского права на контент, содержащийся в документах из множества $\boldsymbol{C}$, можно описать как отображение $\boldsymbol{F}$:

$$\boldsymbol{F}: \boldsymbol{M}\times\boldsymbol{C}\times\boldsymbol{K}\rightarrow \boldsymbol{S} \quad (1)$$

Процесс извлечения $\boldsymbol{M}$ из стеганоконтейнеров $\boldsymbol{S}$ описывается функцией, обратной к $\boldsymbol{F}$:

$$\boldsymbol{F}^{-1}: \boldsymbol{S}\times\boldsymbol{K}\rightarrow \boldsymbol{M}, \boldsymbol{C}. \quad (2)$$

Таким образом стеганографическая система определяется как:

$$\boldsymbol{SF} = (\boldsymbol{SC}, \boldsymbol{C}, \boldsymbol{M}, \boldsymbol{K}, \boldsymbol{S}, \boldsymbol{F}, \boldsymbol{F}^{-1}), \quad (3)$$

где $\boldsymbol{SC}$ – стеганографический канал

$$\boldsymbol{F}: \boldsymbol{C} \rightarrow \boldsymbol{SC}. \quad (4)$$

Множество всех ключей $\boldsymbol{K}$ можно представить семейством множеств $\boldsymbol{K}_\text{г}$ и $\boldsymbol{K}_\text{п}$: $\boldsymbol{K} = \{\boldsymbol{K}_\text{г}, \boldsymbol{K}_\text{п}\}$. Таким образом $\boldsymbol{K}_\text{г}$ − множество ключей для генерации сообщения, $\boldsymbol{K}_\text{г} = \{K_{\text{г}1}, K_{\text{г}2}, \ldots, K_{\text{г}t} \}$, а $\boldsymbol{K}_\text{п}$ − множество ключей для методов внедрения сообщения, $\boldsymbol{K}_\text{п}= \{K_{\text{п}1}, K_{\text{п}2}, \ldots, K_{\text{п}g}\}$.

Ключ $\boldsymbol{K}_\text{г}$ первого рода $\boldsymbol{K}_{\text{г}1}$ ($\boldsymbol{K}_{\text{г}1} \in \boldsymbol{K}_\text{г}$) стеганографической системы является обозначением типа используемого контейнера.

Ключ $\boldsymbol{K}_\text{г}$ второго рода $\boldsymbol{K}_{\text{г}2}$ ($\boldsymbol{K}_{\text{г}2} \in \boldsymbol{K}_\text{г}$) будет использоваться для обозначения стеганографического преобразования.

Следовательно, множество $\boldsymbol{K}_\text{г}$ представляет собой набор $\{\boldsymbol{K}_{\text{г}1}, \boldsymbol{K}_{\text{г}2}\}$.

Ключ $\boldsymbol{K}_\text{п}$ первого рода $\boldsymbol{K}_{\text{п}1}$ ($\boldsymbol{K}_{\text{п}1} \in \boldsymbol{K}_\text{п}$) будет применяться для обозначения множества цветовых каналов необходимых для определения массива пикселей $\boldsymbol{C}$, используемых для стеганографического преобразования.

Ключ $\boldsymbol{K}_\text{п}$ второго рода $\boldsymbol{K}_{\text{п}2}$ ($\boldsymbol{K}_{\text{п}2} \in \boldsymbol{K}_\text{п}$) будет определять базовый пиксель или его значение.

Ключ $\boldsymbol{K}_\text{п}$ третьего рода $\boldsymbol{K}_{\text{п}3}$ ($\boldsymbol{K}_{\text{п}3} \in \boldsymbol{K}_\text{п}$) будет применяться для обозначения множества цветовых каналов, используемых непосредственно для стеганографического преобразования.

Ключ $\boldsymbol{K}_\text{п}$ четвертого рода $\boldsymbol{K}_{\text{п}4}$ ($\boldsymbol{K}_{\text{п}4} \in \boldsymbol{K}_\text{п}$) будет применяться для обозначения параметров для изменения цвета, или других свойств для встраивания информации в $\boldsymbol{C}$.

Таким образом, множество $\boldsymbol{K}_\text{п}$ представляет собой набор $\{\boldsymbol{K}_{\text{п}1}, \boldsymbol{K}_{\text{п}2}, \boldsymbol{K}_{\text{п}3}, \boldsymbol{K}_{\text{п}4}\}$.

Стеганографическая система определяется как:

$$\boldsymbol{SF} = (\boldsymbol{SC}, \boldsymbol{C}, \boldsymbol{M}`, \boldsymbol{K}, \boldsymbol{S}, \boldsymbol{F}, \boldsymbol{F}^{-1}), \quad (5)$$

где $\boldsymbol{K} = \{\boldsymbol{K}_\text{г}, \boldsymbol{K}_\text{п}\}$, $\boldsymbol{K}_\text{г}=\{\boldsymbol{K}_{\text{г}1}, \boldsymbol{K}_{\text{г}2}\}$, $\boldsymbol{K}_\text{п} =\{\boldsymbol{K}_{\text{п}1}, \boldsymbol{K}_{\text{п}2}, \boldsymbol{K}_{\text{п}3}, \boldsymbol{K}_{\text{п}4}\}$.

На основе описанной стеганографической модели был разработан, который использует стеганографическое преобразование для полутоновых пикселей, полученных при растрировании векторных текстовых символов. Оригинальность метода состоит в том, что процессы внедрения / извлечения происходят в пикселях, выбранных на основе соответствия значения цветового канала $K_{п1}$ значению $K_{п2,}$ и изменения цветового значения канала $K_{п3}$ на $K_{п4}$. Непосредственно внедрение / извлечение сообщения происходит при анализе цветового значения канала $K_{п3}$: если $K_{п3}$ пикселя для внедрения является четным, то бит сообщения равен 1, иначе 0. Если при внедрении значение пикселя соответствует необходимому, то изменение не производиться, в ином случае к текущему значению цветового канала $K_{п3}$ добавляется $K_{п4}$. Полный алгоритм реализации метода внедрения сообщения выглядит следующим образом:

Шаг 1. Определение растрового документа-контейнера. Выбор сообщения, которое необходимо скрыть (M_i).

Шаг 2. Добавление к M_i вспомогательной информации (число разрядов длины сообщения, длину сообщения).

Шаг 3. Представление сообщения M_i в двоичном виде.

Шаг 4. Подсчет общего количества знаков n, составляющих M_i.

Шаг 5. Выбор ключевой информации: $K_{п1,}$ $K_{п2}$, $K_{п3}$, $K_{п4}$ ($K_{п4}$%2=1).

Шаг 6. Создание массива пикселей Z, для стеганографического преобразования.

Шаг 7. Определение размера l массива Z.

Шаг 8. Проверка условия: $l \geq n+1$? При выполнении условия – переход к шагу 9, в противном случае – к шагу 5.

Шаг 9. Пока существует i-тый элемент, имеющий значений от 1 до n+1, то выполняются шаги 10-13, в противном случает переход к шагу 14.

Шаг 10. Проверка условия: $K_{п3}(Z_i)$%2=1? При выполнении условия – переход к шагу 11, в противном случае – к шагу 13.

Шаг 11. Проверка условия: символ m_i сообщения равен 0? При выполнении условия – переход к шагу 12, в противном случае – к шагу 9.

Шаг 12. Присвоение $K_{п3}(Z_i)$ нового значения $K_{п3}(Z_i)+ K_{п4}$.

Шаг 13. Проверка условия: символ m_i сообщения равен 1? При выполнении условия – переход к шагу 12, в противном случае – к шагу 9.

Шаг 14. Конец.

Важным этапом данного алгоритма является создание массива пикселей Z, которые будут непосредственно использоваться для стеганографического преобразования. Массив выбирается по проверке на совпадение значения цветового канала $K_{п1}$ каждого пикселя значению $K_{п2}$. В массив добавляется пиксель, находящийся вниз по диагонали от проверяемого. Это позволит получить массив, состоящий из максимально разнородных пикселей, которые не имеют между собой очевидных корреляций. Алгоритм создания массива Z:

Шаг 1. Определения t – ширина документа-контейнера, r – высота документа-контейнера.

Шаг 2. Пока существует j-тый элемент, имеющий значений от 1 до t-1, то выполняются шаги 3-5, в противном случает переход к шагу 6.

Шаг 3. Пока существует n-ый элемент, имеющий значений от 1 до r-1, то выполняются шаги 4-5, в противном случает переход к шагу 2.

Шаг 4. Проверка условия: $K_{п1}(C_{jn}) = K_{п2}$? При выполнении условия – переход к шагу 5, в противном случае – к шагу 3.

Шаг 5. Поместить пиксель C_{j+1n+1} в массив Z.

Шаг 6. Конец.

Алгоритм извлечения сообщения является обратным к алгоритму внедрения.

Алгоритмы внедрения и извлечения сообщения характеризуются линейной сложностью: $O(n)$.

Данный метод может быть использован для скрытия информации в текстовых документах, представленных в формате растровой графики. Эффективность метода зависит от особенностей изображения-контейнера, таких как количество пикселей с сопоставимыми значениями в одном или нескольких цветовых каналах. Кроме того, этот метод можно применять и к изображениям с незначительным сжатием, просто увеличивая значение параметра $K_{п4}$. Тем не менее, в таком случае внедрение данных будет более заметным.

Библиографические ссылки

1. Шутько Н. П., Листопад Н. И., Урбанович П. П. Моделирование стеганографической системы в задачах по охране авторских прав // 8-я МНТК Информационные технологии в промышленности, ITI-2015. Тезисы докладов, Минск, 2-3.04.2015. Минск: ОИПИ НАНБ, 2015. С. 30-31.

2. Быканова А. С. Методы распознавания математических формул в электронных документах // Актуальные проблемы авиации и космонавтики: сб. материалов XIV Междунар. науч.-практ. конф., посвящ. Дню космонавтики (09-13 апреля 2018 г., Красноярск): в 3 т. Т. 2. Красноярск: СибГУ им. М. Ф. Решетнева, 2018. Т. 2. №. 14. С. 133-134.

3. Агеев В. Н., Соломыков В. С. Моделирование процесса растрирования векторных шрифтов в выводных устройствах низкого разрешения // Известия Тульского государственного университета. Технические науки. Тула: ТулГУ, 2013. №. 3. С. 9-16.

4. Савельева М. Г., Урбанович П. П. Метод стеганографического преобразования web-документов на основе растровой графики и модели RGB // Труды БГТУ. Серия 3: Физико-математические науки и информатика. Минск: БГТУ, 2022. №. 2 (260). С. 99-107.

5. Стеганография, цифровые водяные знаки и стеганоанализ. / В. Г. Грибунин [и др.]: Монография. М.: Вузовская книга. 2009. 217 с.

6. 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. Серия 3: Физико-математические науки и информатика. Минск: БГТУ, 2015. №. 6 (179). С. 152-156.