

## СТАТИСТИЧЕСКАЯ ОБРАБОТКА РЕЗУЛЬТАТОВ РАСТРИРОВАНИЯ WEB-ДОКУМЕНТОВ

Электронный документ-контейнер, защищаемый от несанкционированного использования или модификации с помощью стеганографии, может изначально создаваться как на основе растровой, так и векторной графики [1]. Оригинальный контент при этом может преобразован из форматов одной графики в другую. При конвертации или преобразовании текстовых документов-контейнеров одним из наиболее важных вопросов является растривание текста: контуры букв начинают расплываться, цвет по контуру переходит в градиент. Однако эту ситуацию можно использовать для встраивания секретной информации (цифрового водяного знака, ЦВЗ) в защищаемый контент.

Для увеличения пропускной способности стеганоканала, т. е. для внедрения большего объема информации в расчете на единицу объема контейнера, следует в элементах переходных оттенков растриванных символов выделить преобладающие оттенки.

Ранее было проанализировано 50 страниц стандартизированного оформления текстовых документов в формате PNG (преобразованных из PDF): 8 содержали элементы, которые либо обязательно преобразуются в растр при конвертации в PDF (схемы, графики, фигуры), либо уже являются таковыми (рисунки); 17 страниц содержали различные элементы сложного оформления (формулы, таблицы, листинги), которые могут стать растровыми; 25 – содержали только текст. Перечисленные источники разделены условно на 4 группы: 1) все страницы, 2) страницы только с текстом, 3) страницы с элементами сложного оформления, 4) страницы с растровыми элементами.

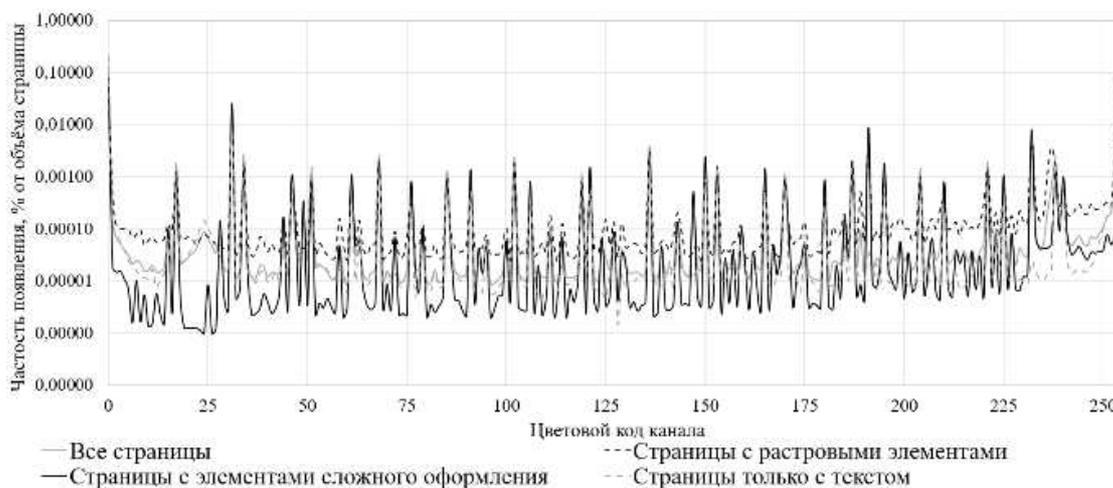
На рис. 1–3 представлены полученные усредненные результаты появления пикселей определенного цветового оттенка, соответствующего коду от 0 до 255 для каждого канала модели RGB.

Следует отметить, что на всех проанализированных страницах объем информации, представленной в текстовом виде, превышает объем иной информации. На большинстве анализируемых страниц преобладают черно-белые оттенки, за исключением нескольких, имеющих небольшие (до 1% от всего объема) вкрапления иного цвета (рамка, стрелка и т. д.).

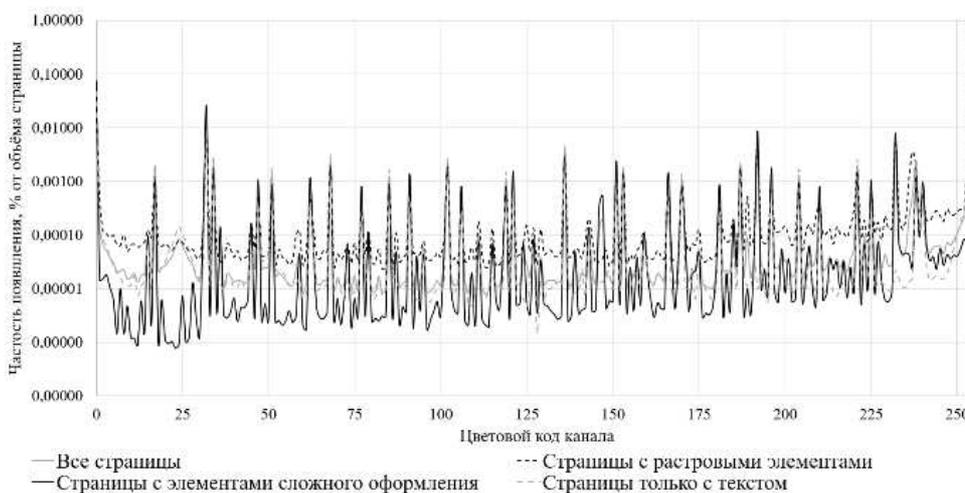
Для красного канала R (рис. 1) наибольшие относительные частности отмечены для кодов, соответствующего следующим десяти-



**Рисунок 1 – Распределение пикселей по цветовым оттенкам для канала R**



**Рисунок 2 – Распределение пикселей по цветовым оттенкам для канала G**



**Рисунок 3 – Распределение пикселей по цветовым оттенкам для канала B**

ным значениям данного цвета: 0, 17, 34, 51, 68, 85, 102, 119, 136, 153, 170, 187, 204, 221, 238, 255, несколько меньшие – для следующим кодов: 64, 79, 94, 123, 152, 167, 192, 196, 226, 233, 236, 237, 253, 254.

Для канала G (рис. 2) соответственно: 0, 17, 34, 51, 68, 85, 102, 119, 136, 153, 170, 187, 204, 221, 238, 255 и 31, 46, 61, 76, 91, 106, 121, 150, 165, 180, 191, 195, 210, 225, 232, 236, 237, 253, 254.

Для канала B (рис. 3) соответственно: 0, 17, 34, 51, 68, 85, 102, 119, 136, 153, 170, 187, 204, 221, 238, 255 и 32, 46, 62, 77, 91, 106, 121, 147, 151, 166, 181, 192, 196, 210, 225, 232, 236, 237, 253, 254.

Пиковые значения для первой из двух рассматриваемых групп (0, 17, 34, 51, 68, 85, 102, 119, 136, 153, 170, 187, 204, 221, 238, 255) во всех распределениях соответствуют цвету от черного (0) до белого (255) для 16 различных оттенков: 0 0 0; 17 17 17; ...; 255 255 255). Так как белый цвет – это фон, то оставшиеся 15 оттенков соответствуют отображениям элементов буквенных символов. Такое количество оттенков необходимо для отображения всех растрованных букв с учетом особенностей их строения.

Для извлечения из полученных результатов максимальной пользы их необходимо статистически обработать. Математическая статистика решает следующие основные задачи [2]:

- статистическое оценивание неизвестных параметров распределений;
- проверка статистических гипотез;
- установление статистических зависимостей между случайными величинами.

В данном эксперименте применены первичные методы (методы, с помощью которых можно получить показатели, непосредственно отражающие результаты производимых в эксперименте измерений): математическое ожидание (1), дисперсия (2) и среднее квадратичное отклонение (3) [2]. Для анализа использованы все полученные результаты цветовых каналов кроме значения 0 (0 0 0 – цвет фона).

$$M\xi = \sum_{i=1}^n x_i p_i ; \quad (1)$$

$$D\xi = M(\xi^2) = M(\xi - M\xi)^2 ; \quad (2)$$

$$\sigma = \sqrt{D\xi} . \quad (3)$$

После статистического анализа получены следующие результаты:

- для канала R  $M\xi = 8,4015$ ,  $D\xi = 1423,0861$ ,  $\sigma = 37,7238$ ;
- для канала G  $M\xi = 8,3006$ ,  $D\xi = 1408,7782$ ,  $\sigma = 37,5237$ ;

– для канала  $B M\xi = 8,6334$ ,  $D\xi = 1458,1356$ ,  $\sigma = 38,1855$ .

Результаты каждого канала подобным остальным, что означает примерно равно распределение значений в цветовых каналах. Но в то время значения не совпадают, что означает наличие не только чистых оттенков градации серого (0 0 0; 17 17 17; ...; 255 255 255).

Выбор соответствующего цветового оттенка для размещения тайной информации (ЦВЗ) позволяет повышать пропускную способность тайного стеганографического канала, снижать эффективность визуальных или иных атак на стеганоконтейнер.

#### ЛИТЕРАТУРА

1. Урбанович, П. П. Защита информации методами криптографии, стеганографии и обфускации: учеб.-метод. пособие / П. П. Урбанович. – Минск: БГТУ, 2016. – 220 с.

2. Ковалев, Е. А. Теория вероятностей и математическая статистика для экономистов : учебник и практикум для прикладного бакалавриата / Е. А. Ковалев, Г. А. Медведев ; под общ. ред. Г. А. Медведева. – М. : Издательство Юрайт, 2016. – 284 с.

УДК 004.56+003.26

Асп. М.Г. Савельева  
(БГТУ, г. Минск)

### **ИССЛЕДОВАНИЕ РАСТРИРОВАНИЯ ГРАФЕМ ДЛЯ ЗАЩИТЫ АВТОРСКОГО ПРАВА НА ОСНОВЕ СТЕГАНОГРАФИЧЕСКИХ ПРЕОБРАЗОВАНИЙ**

Электронные текстовые документы очень легко меняют свое исходное состояние при прохождении через Интернет или преднамеренно трансформируются и видоизменяются. Любой из них можно считать изображением и частью растровой или векторной графики. При преобразовании или переводе текстового документа-контейнера  $S$  одним из ключевых вопросов является растривание текста. Однако растривание можно использовать как преимущество для встраивания конфиденциальной информации (ЦВЗ) в контейнеры.

Приняты определенные правила форматирования и оформления текстовых документов. В общем случае их можно записать так: для печатных документов – шрифт Times New Roman; для документов, обрабатываемых преимущественно в электронном виде, возможно также применение шрифтов Arial, Helvetica, Verdana; размер 12–14 пт – для основного текста; 10–14 пт – для таблиц и подписей.