О ПРИМЕНЕНИИ СТАТИСТИЧЕСКИХ МЕТОДОВ В СТЕГАНОАНАЛИЗЕ ГРАФИЧЕСКИХ ИЗОБРАЖЕНИЙ

Лобач С. В., Меркулов Р. И., Акулич В. Н.
Кафедра математического моделирования и анализа данных,
Белорусский государственный университет
Минск, Республика Беларусь
E-mail: lobachS@bsu.by, {merkylovecom, fizik2009}@mail.ru

В статье рассматриваются методы стеганографии, встраивание информации в изображения, проводится сравнительный анализ статистических методов стеганоанализа.

Введение

Основная цель стегоанализа [1] состоит в том, чтобы обнаружить факт наличия скрытого сообщения в модифицированном контейнере: аудиофайле, графическом изображении, видеопоследовательности (пассивный стегоанализ) -– и, если возможно, извлечь, вскрыть, подменить или уничтожить скрытое сообщение (активный стегоанализ). В стегоанализе можно разработать эффективный алгоритм обнаружения и извлечения сообщения из модифицированного контейнера в том случае, когда известен алгоритм встраивания информации. Но, как правило, алгоритм сокрытия сообщения в контейнере неизвестен. Поэтому основная задача состоит в разработке таких алгоритмов, которые были бы достаточно эффективны при проведении стегоанализа для определенного семейства алгоритмов сокрытия информации.

Можно выделить несколько основных направлений развития стегоанализа [1]:

- стегоанализ на основе теории статистического распознавания образов;
- параметрический статистический стегоанализ;
- стегоанализ, основанный на идентификации скрытого сообщения.

Первое направление (известное как слепой стегоанализ) основано на использовании статистических методов классификации с обучением.

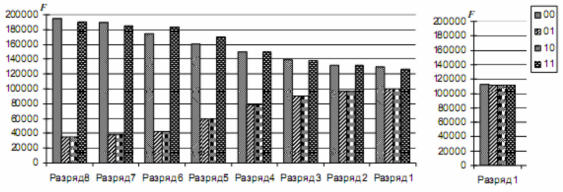
Второе направление стегоанализа предполагает известные параметрические модели, которыми описываются контейнер, модифицированный контейнер и скрываемое сообщение.

В третьем направлении проблема стегоанализа рассматривается как проблема идентификации. относительно контейнера и скрываемого сообщения предполагается, что они независимы.

I. Статистические методы стеганоанализа

В данной статье рассматривается задача пассивного стеганоанализа, т.е. задача обнаружения самого факта наличия скрываемого сообщения в модифицированном контейнере. Применяются статистические методы анализа бинарных последовательностей, которые описывают исходные изображения. Проводятся компьютерные эксперименты на реальных изображениях, в которых имеются скрытые сообщения.

Оценка числа переходов значений младших бит. Метод основывается на том факте, что между младшими битами соседних элементов, а также между ними и остальными битами в естественных контейнерах имеются корреляционные связи [2]. Так как последовательности битов являются двоичными последовательностями, то анализируется четыре вида переходо: $0 \to 0$, $0 \to 1$, $1 \to 0$, $1 \to 1$. По полученным результатам строится гистограмма, где каждый столбец соответствует одному из переходов. Приведем пример построения такой гистограммы.



а б

Рис. 1 – Гистограмма частот переходов битовых значений: а – пустого контейнера, б – стегоконтейнера

Для пустого и модифицированного контейнера число переходов в потоке НЗБ (наименее значимых бит) будет разным. Распределение НЗБ модифицированного контейнера имеет, как правило, случайный характер. Пустому контейнеру не свойственно примерно одинаковое число переходов в потоке НЗБ для всех состояний.

Оценка частот появления битовых серий. Метод позволяет оценить равномерность распределения элементов в исследуемой последовательности на основе анализа частоты появления нулей и единиц в серии из k бит [3]. В битовом представлении исследуемой последовательности подсчитывается, сколько раз встречаются нули и единицы ($k = 1$), серии-двойки 00, 01, 10, 11 ($k = 2$), серии-тройки 000, 001, 010, 100, 101, 110, 111 ($k = 3$). На основании этого строится гистограмма. Для незаполненных контейнеров не является характерным, чтобы значения частот всех компонентов находились достаточно близко. При внедрении информации значения частот сближаются. Этот факт используется при стеганоанализе.

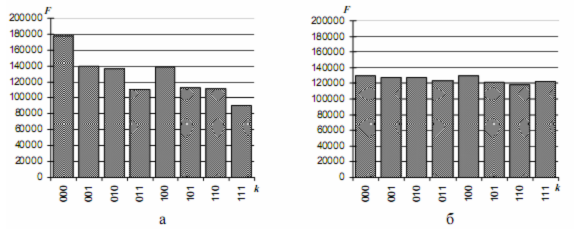


Рис. 2 – Гистограмма частот серии-тройки ($k = 3$) в потоке НЗБ: а – пустого контейнера, б – стегоконтейнера

Для незаполненных JPEG-изображений не является характерным, чтобы значения частот всех компонентов находились достаточно близко (рис. 2 а). При внедрении информации значения частот сближаются (рис. 2 б). Этот факт используется при анализе. Результаты работы метода зависят от стеганографического преобразования, используемого для встраивания скрываемых данных, а также от их объема. Как правило, выявление факта скрытия осуществимо при заполнении контейнера на 60% и выше.

Анализ распределения значений на основе критерия хи-квадрат. В методе используется анализ гистограммы, полученной по элементам изображения и оценка распределения пар значений этой гистограммы [3, 4].

Метод хи-квадрат является универсальным, так как подходит для анализа изображений, созданных различными программами скрытия. Однако результаты работы метода по критерию хи-квадрат в значительной мере зависят от способа скрытия данных. При последовательной записи в НЗБ элементов контейнера метод обеспечивает хорошие результаты (рис. 3), а при псевдослучайном выборе младших бит и рассеивании сообщения по всей длине контейнера метод не срабатывает. Кроме того, существует возможность выбора отдельных областей изображения для их последующего анализа. Такой подход позволяет выявлять наличие информации, скрытой псевдослучайным образом.

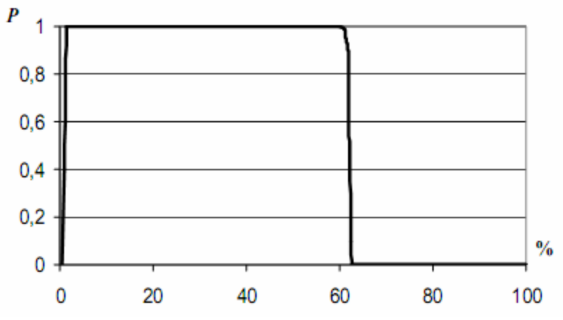


Рис. 3 – Вероятность встраивания по критерию χ^2 при анализе стегоконтейнера, полученного методом последовательной замены

Анализ гистограмм частот элементов изображения. Метод позволяет оценить равномерность распределения элементов анализируемого изображения, а также определить частоту появления конкретного элемента [5]. Если разброс частот появления элементов в цветовых составляющих изображения стремится к нулю, то контейнер содержит скрытые данные. В противном случае контейнер считается пустым. Для изображений в JPEG-формате строится гистограмма частот квантованных дискретных косинусных коэффициентов. Экспериментально обнаружено, что огибающая гистограммы пустого изображения имеет более гладкий характер (рис. 4 а) по сравнению с гистограммами изображений, содержащими стеганографическое вложение (рис. 4 б).При значительных объемах скрываемой информации гистограммы часто приобретают ступенчатый характер, что нетипично для обычных JPEG-изображений.

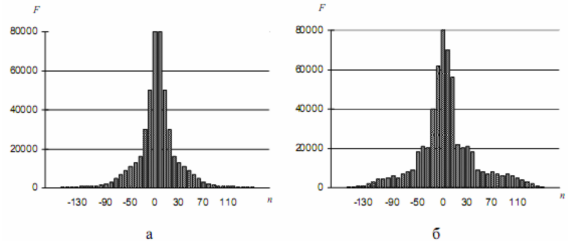


Рис. 4 – Гистограмма частот дискретных косинусных коэффициентов: а – исходного изображения, б – изображения, содержащего скрытую информацию

Заключение

В статье приведен краткий обзор стеганографических методов, рассмотрены некоторые статистические методы, применяемые в стеганографии. Статистические методы не являются средством, позволяющим со 100% надежностью определять наличие скрытой информации. Они дают возможность аналитику с определенной вероятностью судить о том, используется стеганография или нет. Проведенные компьютерные эксперименты не позволяют выделить тот или иной статистический метод, их надо проводить в совокупности.

Список литературы

1. Харин, Ю. С. Стеганографические методы защиты информации: обзор / Ю. С. Харин, М. С. Абрамович // Управление защитой информации. – 2009. – Т. 13, № 1. – С. 58–64.
2. Швадченко, И. В. Методы стеганоанализа для графических файлов / И. В. Швадченко // Киев, 2010.
3. Основы компьютерной стеганографии / А. В. Аграновский [и др.]. – М.: Радио и связь, 2003.
4. Варновский, И. П. Математика и безопасность информационных технологий в МГУ / И. П. Варновский, О. А. Голубев, О. А. Логачев // Современные направления стеганографии. – Издательство: МЦНО, 2005. – С. 32–64.
5. Fridrich, J. On steganographic embedding efficiency / J. Fridrich, P. Lisonek, D. Soukal // Lecture notes in Computer Science. – New York, 2006. – Vol. 4437. – P. 123–132.