УДК 621.391

М.Ю. Жилкин

Стегоанализ графических данных в различных форматах

Предлагается метод определения присутствия скрытой информации в стеганографических системах, базирующихся на изменении наименее значащих битов (LSB). Экспериментально показано, что метод имеет достаточно высокую эффективность.

Введение

Стеганография – это наука о сокрытии факта передачи информации, которая уходит корнями в древние времена. В настоящее время в связи с бурным развитием вычислительной техники и новых каналов передачи информации появились новые стеганографические методы, в основе которых лежат особенности представления информации в компьютерных файлах, вычислительных сетях и т.п. Методы современной компьютерной стеганографии находят применение в области военной и правительственной связи, защиты авторских прав, для решения задач обеспечения информационной безопасности. Актуальность проблемы информационной безопасности постоянно растет и стимулирует разработку новых методов стеганографии и атак на уже существующие методы.

Цель работы – построение эффективного метода «автоматического» (т.е. без участия человека) определения факта наличия скрытых данных в графических файлах форматов BMP и JPEG. Предлагается новый метод, основанный на применении сжатия данных. Среди достоинств метода – высокая скорость, эффективность, применимость к широкому классу методов LSB.

Постановка задачи и описание метода

Распространенные на сегодня стеганографические пакеты используют различные методы включения данных в графические файлы. Скрытое сообщение может располагаться в младших битах последовательно, последовательно с заполнением остатка случайными данными или «разбросанным» методом – байты контейнера выбираются случайно.

Труднее всего подвергаются стеганографическому анализу контейнеры, в которые «скрытая» информация добавляется «разбросанным» методом. Известные в литературе атаки дают результаты при заполнении изображения не менее чем на 50% от его емкости [2].

Включение данных в BMP-файлы производится без всяких дополнительных преобразований. Самым простым и поэтому наиболее распространенным форматом этого класса является 24-битный BMP. Описываемый метод работает только с изображениями этого формата.

Алгоритм сжатия JPEG состоит из 4 основных этапов: преобразования цветового пространства, дискретного косинусного преобразования (ДКП), квантования и неискажающего сжатия по алгоритму Хаффмана. Скрытые данные включаются перед последним этапом алгоритма.

В работе описывается новый класс алгоритмов стегоанализа, базирующихся на сжатии данных. В качестве инструмента могут выступать широко распространенные программы-архиваторы. Идея метода состоит в следующем: поток случайных данных сжимается хуже, чем поток, где встречаются повторяющиеся последовательности. Информация, включаемая в младшие биты контейнера, как правило, предварительно шифруется и, возможно, сжимается, поэтому является псевдослучайной. Степень сжатия контейнеров используется для определения наличия в них скрытой информации.

Перейдем к формальному описанию алгоритма. Пусть $X = \{x_1, \ldots, x_N\}$ – последовательность байтов в поле данных изображения BMP, где $|X| = N$ - длина последовательности. Разобьем последовательность X на d равных отрезков и обозначим каждый отрезок X_i, где $i = 1, 2, \ldots, d$. Пусть $\psi(X)$ - алгоритм сжатия, примененный к последовательности X.

Тогда обозначим за

$$f(X,n) = \frac{|\psi(X_n)|}{|X_n|}$$

коэффициент сжатия отрезка n последовательности X алгоритмом ψ. Обозначим за $\varphi(X)$ псевдослучайное изменение младших битов всех байтов последовательности X.

Пусть X – последовательность, которая подается на вход программе, а $Y = \varphi(X)$ – полученная из нее новая последовательность. Исходная последовательность X сжимается сильнее по сравнению с измененной последовательностью Y.

Введем новую величину $\delta(X,n) = |f(X,n) - f(Y,n)|$. Те отрезки последовательности X, которые не содержали «скрытую» информацию, сжимаются лучше, чем соответствующие им отрезки последовательности Y, и, напротив, коэффициенты сжатия отрезка последовательности X со «спрятанной» информацией и отвечающего ему отрезка последовательности Y отличаются незначительно. Для определения факта включения информации выбирается пороговое значение для величины δ и производится оценка количества отрезков, на которых значение величины не превышает порог.

Экспериментальный анализ

Для экспериментального исследования метода была подготовлена серия из 140 изображений («контейнеров») форматов 24-bit BMP и JPEG разного разрешения и качественного содержания. На вход программы подавались исходные контейнеры, контейнеры, заполненные на 10%, 20%, и т.д. от своей емкости. Были проверены различные методы заполнения контейнеров.

На рис. 1–2 приведены данные по работе алгоритма с архиватором ZIP для BMP и встроенным алгоритмом сжатия для JPEG.

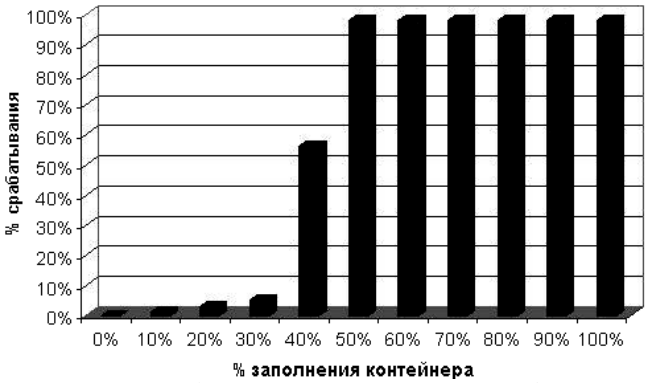

Рис. 1. Результат работы алгоритма на контейнерах формата BMP

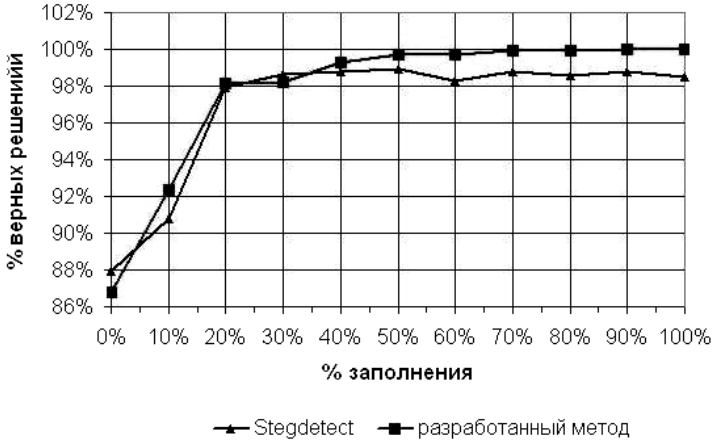

Рис. 2. Результат работы алгоритма на контейнерах формата JPEG в сравнии с стандартной утилитой StegDetect

Таким образом, предлагаемый метод позволяет надежно без участия человека выявлять факт включения данных в изображение. Ошибка при заполнении контейнера на 40% и более не превышает 2%.

Литература

1. Ryabko B., Astola J. Universal Codes as a Basis for Time Series Testing // Statistical Methodology. 2006. – Vol. 3. P. 375–397.

2. Westfeld A., Pfitzmann A. Attacks on Steganographic Systems. Breaking the Steganographic Utilities EzStego, Jsteg, Steganos and S-Tools – and Some Lessons Learned. Lecture Notes in Computer Science, 1768:61–75, 2000

3. Provos N., Honeyman P. Hide and Seek: An Introduction to Steganography. IEEE Security & Privacy, may/june 2003, pp. 32-44.

4. Dabeer O., Sullivan K., Madhow U., Chandrasekaran S., and Manjunath B.S. Detection of hiding in the least significant bit. In IEEE Trans. on Signal Processing,volume 52, pp. 3046–3058, Oct. 2004.

5. Fridrich J., Goljan M., and Du R. Reliable Detection of LSB Steganography in Color and Grayscale Images. Proc. of the ACM Workshop on Multimedia and Security, pp. 27–30, 2001.

Жилкин Михаил Юрьевич
ГОУ ВПО Сибирский государственный университет телекоммуникаций и информатик» (СибГУТИ), аспирант
Эл. почта: fionov@neic.nsk.su.

M.U. Djilkin

Stegoanalyze graphic data in various formats

The method of definition of presence of the latent information in stenographic the systems which are based change least meaning bitsIs offered. It is experimentally shown, that the method has high enough efficiency.