Таблица 4. Статистика тестирования в режиме определения типа атак (число выходных нейронов равняется 23)

служба	определенные атаки	ложное срабатывание	распознанные атаки
auth	108(100%)	0	108(100%)
domain	113(100%)	0	113(100%)
eco_I	1252(99,9%)	0	1239(99,0%)
ecr_I	281034(99,9%)	13(3,77%)	281034(100%)
finger	186(92,1%)	10(2,14%)	185(99,5%)
ftp	418(98,3%)	26(6,97%)	418(100%)
ftp_data	856(92,7%)	31(0,82%)	636(74,3%)
http	2400(99,7%)	96(0,16%)	2400(100%)
IRC	1(100%)	1(2,38%)	1(100%)
pop_3	123(100%)	0	123(100%)
smtp	122(97,6%)	35(0,36%)	119(97,5%)
telnet	284(96,6%)	15(6,85%)	272(95,7%)

Таблица 5. Статистика тестирования модели 4 в режиме классификации атак (около 30 сервисов)

класс	всего	обнаружено	распознано
DoS	286369	286369(100%)	286295(99,9%)
U2R	49	33(67,3%)	32(97,0%)
R2L	1119	442(39,5%)	427(96,6%)
Prode	1320	1311(99,3%)	1288(98,2%)
нормальное состояние			
normal	83281	---	77673(93,2%)

Сравнение результатов опытов с разным количеством выходных нейронных элементов показало, что полученные данные практически идентичны (см. таблица 3 и таблица 4).

Оценим эффективность модели 1 в сравнении с моделью 4 (см. таблица 2 и таблица 5). В нашем эксперименте параметры входного вектора делились на три группы - те, что принимают целочисленные значения; ключи 0/1; параметры, величины которых лежат в диапазоне от 0 до 1. Далее каждая группа подавалась на обработку соответствующей RNN. Расчет был сделан на улучшение качества обучения за счет того, что каждая RNN в этом случае обрабатывает однородные по структуре данные. Особенностью модели 4 является ее пониженная чувствительность к атакам класса U2R и R2L. В то же время эта модель может успешно применяться для распознавания наиболее широко представленных в базе KDD-99 атак класса DoS. Еще одним преимуществом модели 4 можно считать скорость ее обучения. За счет уменьшения количества связей в модуле вычисления главных компонент снижается нагрузка на системные ресурсы в процессе пересчета сети.

5. ЗАКЛЮЧЕНИЕ

В работе рассмотрены различные подходы к построению систем обнаружения атак, которые базируются на нейросетевых технологиях. Путем комбинирования двух различных нейронных сетей, а именно RNN и MLP, можно идентифицировать и распознавать атаки на компьютерные сети с достаточно высокой степенью точности. В качестве базы данных для тестирования предложенных методов использовалась база KDD-99. Основными преимуществами использования подходов, основанных на нейронных сетях, является способность адаптироваться к динамическим условиям и быстрота функционирования, что особенно важно при работе системы в режиме реального времени.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. 1999 KDD Cup Competition. http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html
2. Головко В.А. Нейронные сети: обучение, организация и применение. Кн. 4: Учеб. пособие для вузов / Общая ред. А.И. Галушкина. – М.: ИПРЖР, 2001. – 256 с.
3. Hyvaerinen A., Oja E. Independent component analysis: algorithms and applications // Neural Networks, №13, 2000, - P. 411-430.

УДК 681.3

Горбашко Л.А.

МЕТОД СКРЫТИЯ ИЗОБРАЖЕНИЙ С ПРИМЕНЕНИЕМ ВЕКТОРНОГО КВАНТОВАНИЯ

ВВЕДЕНИЕ

Задача защиты информации от несанкционированного доступа приобретает все большую актуальность в современном мире. Развитие информационных технологий дало новый толчок для развития компьютерной стеганографии - скрытую передачу данных в маскирующем сигнале. Сообщения встраивают в цифровые данные, имеющие аналоговую природу (речь, аудиозаписи, изображения, видео).

Для скрытой передачи информации выбирается контейнер, которым чаще всего служит графический файл. В него

***Горбашко Лариса Ашотовна**, ст. преподаватель каф. интеллектуальных информационных технологий Брестского государственного технического университета.*
Беларусь, БрГТУ, 224017, Беларусь, г. Брест, ул. Московская, 267, e-mail: lagorbashko@bstu.by.

особым образом встраивается сообщение, которое нужно передать тайно. В результате получается комбинированное изображение – стего, которое и передается по каналу связи. Одним из основных требований при встраивании данных является соблюдение прозрачности (отсутствии видимых искажений) стего.

В настоящее время для встраивания данных чаще всего используются прямые методы (встраивание в пространственной области) и частотные методы (встраивание в частотной области). Прямые методы не дают достаточной робастности (устойчивости к преобразованиям стего и атакам на стего, которым подвергается любой сигнал при передаче по каналу связи). Поэтому современные исследования направлены на встраивание информации в частотной области. При этом перед встраиванием данных производят разложение контейнера на любые частотные составляющие (преобразования Фурье, вейвлет-, дискретное косинусное и т.п.). При этом основным показателем робастности метода является его устойчивость к сжатию информации. Для обеспечения робастности стего к сжатию следует использовать преобразование, заложенное в алгоритме сжатия. Так, метод с использованием дискретного косинусного преобразования (ДКП) будет устойчив к JPEG-сжатию [1]; если же мы хотим добиться устойчивости к сжатию по стандарту JPEG2000, то при разложении контейнера на частотные составляющие следует использовать вейвлет-преобразование.

Кроме того, последние исследования направлены на увеличение объема встроенной информации, что вступает в противоречие с требованиями робастности и прозрачности. Метод считается приемлемым, если он позволяет робастно встраивать 1% информации по отношению к объему контейнера [2].

Анализируя подобные работы, можно привести современные значения объема встраиваемых данных. В работе [3] в качестве контейнера используется видео-, а в качестве сообщения- видео- или аудио- файлы. Сообщение внедряется в коэффициенты ДКП, 2 бита на блок 8x8 точек, что составляет 0.4 процента объема контейнера. В работе [4] встраивание производится с использованием квантования и объем внедренных данных составляет 1%. Наибольший процент внедрения получен в предлагаемом авторами [1] методе – до 25%, при этом используется ДКП контейнера.

Если ставить задачу робастности стего к сжатию JPEG2000 как перспективного стандарта, то следует анализировать методы встраивания с использованием вейвлет- преобразования. В работе [5] основная задача состоит в возможности минимального количества внедренных данных для создания цифровых водяных знаков. Внедренный объем данных должен быть достаточным, чтобы не нарушить корреляционные характеристики контейнера. При использовании вейвлетов Хаара и встраивании в детализируюшие коэффициенты контейнера анализируется алгоритм Тиана и его предложенная модификация, при этом сообщение составляет приблизительно 10% от объема контейнера, минимальный объем-4%. Емкость сообщения, внедренного по методу [6], составляет 4-10%. Внедрение производится в незначимые аппроксимирующие вейвлет- коэффициенты.

В данной статье предлагается метод внедрения изображения, объем которого составляет 10-25% от объема контейнера. Внедрение происходит в частотной области с использованием вейвлет- преобразования. Отличительной особенностью данного метода является векторное квантование сообщения, что позволяет увеличить объем встроенных данных.

ВЕКТОРНОЕ КВАНТОВАНИЕ

Квантование является одним из способов сокращения объема информации. Когда набор значений сигнала квантуется совместно как единый вектор, такой процесс называется *векторным квантованием* (VQ- vector quantization). При векторном квантовании весь набор отсчетов сигнала разбивается на векторы, векторы объединяются в группы (кластеры). Каждый кластер представлен одним вектором, т.е. каждый n-мерный вектор X изображения заменяется на n-мерный вектор Y, представляющий данный кластер. Набор векторов Y называется *кодовой книгой (codebook)*. После построения кодовой книги проводят квантование изображения. В линию связи передаются только индексы векторов (кодированное изображение). Схематически процесс прямого векторного квантования изображен на рисунке 1. При восстановлении изображения каждый индекс заменяется на вектор из кодовой книги с соответствующим номером. Поскольку вектор не точно соответствует исходному, при квантовании неизбежно возникают искажения. Степень искажений зависит от величины кодовой книги.

Некоторые кодовые книги рассчитываются заранее и не изменяются – это фиксированные кодовые книги. Другие кодовые книги могут обновляться в процессе работы – это адаптивные кодовые книги. Построение оптимальной кодовой книги является достаточно сложной задачей. Кодовая книга должна быть известна получателю сообщения и обычно зависит от данных [7].

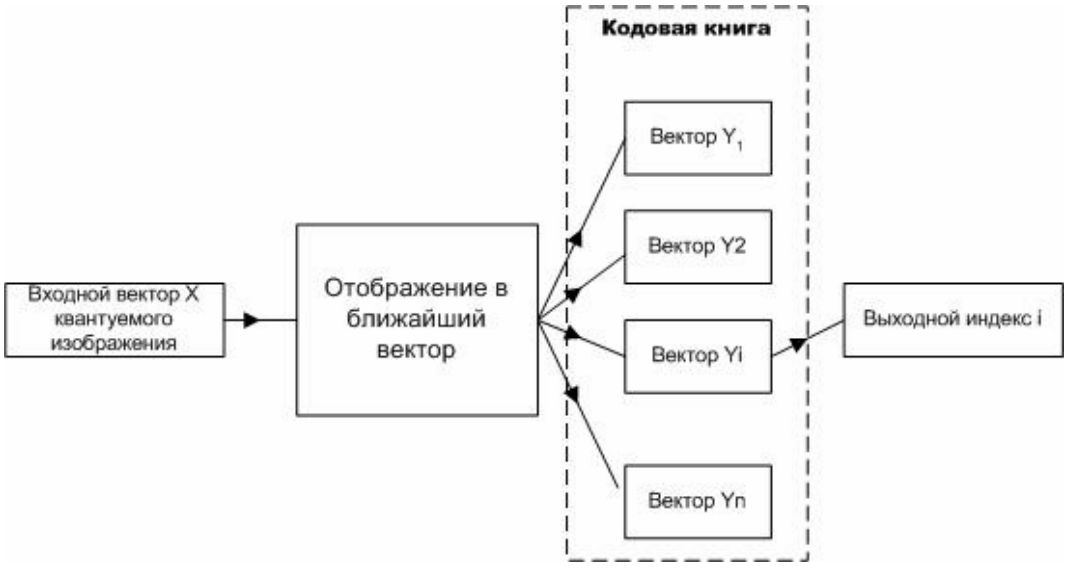

Рис. 1. Прямое векторное квантование изображения.

ВЕЙВЛЕТ- ПРЕОБРАЗОВАНИЕ ХААРА

Вейвлет- преобразование заключается в замене изображения средним значением его пикселов [7], при этом последовательность, например, $\{a_0, a_1, a_2, a_3\}$ заменяется последовательностью $\{a_0, a_1, d_0, d_1\}$. При использовании вейвлетов Хаара преобразование проводится по формулам:

$$a_{k,j} = \frac{a_{k-1,j} + a_{k-1,j+1}}{2} * \sqrt{2^k};$$

$$d_{k,j} = \frac{a_{k-1,j} - a_{k-1,j+1}}{2} * \sqrt{2^k},$$

где a - коэффициенты аппроксимации;

d - коэффициенты детализации;

k - уровень преобразования;

j – номер отсчета сигнала.

Коэффициентами 0-го уровня преобразования являются дискретные отсчеты изображения. При проведении одного уровня преобразования сигнала множество отсчетов делится на две половины: низкочастотная (L), содержащая коэффициенты аппроксимации (полусуммы отсчетов изображения), и высокочастотная (H), содержащая коэффициенты детализации (полуразности). По имеющимся отсчетам всегда можно восстановить исходное изображение. Можно повторять преобразования с полученной последовательностью – это многоуровневое вейвлет- преобразование. При этом количество аппроксимирующих коэффициентов будет уменьшаться в 2 раза на каждом шаге.

Это так называемое одномерное преобразование. Двумерное преобразование одного уровня проводится в 2 этапа: сначала строки матрицы изображения, затем столбцы уже преобразованной матрицы.

АЛГОРИТМ ВНЕДРЕНИЯ И ИЗВЛЕЧЕНИЯ ИЗОБРАЖЕНИЯ

Предлагается следующий алгоритм для внедрения изображения- сообщения в изображение- контейнер:

1. Дискретные отсчеты контейнера подвергаются двумерному одноуровневому вейвлет- преобразованию Хаара.
2. Строим кодовую книгу. Для этого матрица отсчетов сообщения разбивается на блоки 8x8 точек, будем называть эти блоки векторами X. Из векторов составляется кодовая книга Y, в которую включается каждый вектор, представляющий новый кластер. Для открытия нового кластера выбирается значение евклидова расстояния E, при котором искажения сообщения допустимы: $E = \sqrt{\sum_{i=1}^{8}\sum_{j=1}^{8}(x_{ij} - y_{ij})^2}$. Если величина E между рассматриваемым вектором X и всеми существующими векторами Y в кодовой книге больше заданной, то вектор X добавляется в кодовую книгу.
3. Производим векторное квантование сообщения. Для этого для каждого вектора производится поиск ближайшего по евклидову расстоянию вектора в кодовой книге, и вектор сообщения заменяется индексом найденного вектора в кодовой книге. Таким образом уменьшается объем встраиваемых данных – вместо 64 отсчетов сигнала сообщение представлено одним индексом. Максимальная величина индекса зависит от размера кодовой книги.
4. Индексы векторов встраиваются в вейвлет- коэффициенты аппроксимации контейнера по аддитивному алгоритму [2], коэффициенты детализации не меняются.
5. Производим обратное вейвлет- преобразование и получаем стего.

Извлечение сообщения происходит при известном контейнере. Кроме того, должна быть известна кодовая книга. Ее можно передавать по другому каналу связи либо использовать фиксированную кодовую книгу. Последовательность действий при извлечении сообщения обратна последовательности встраивания.

РЕЗУЛЬТАТЫ ЭКСПЕРИМЕНТОВ

В качестве контейнера и сообщения использовались графические файлы формата .BMP и .JPG с 24-битной градацией цвета: контейнер размером 256x256, сообщение размером до 128x128. Для встраивания и извлечения сообщений была разработана программа на языке пакета MatLab 6.5. Кодовая книга строилась с евклидовым расстоянием E=128, т.к. боль-

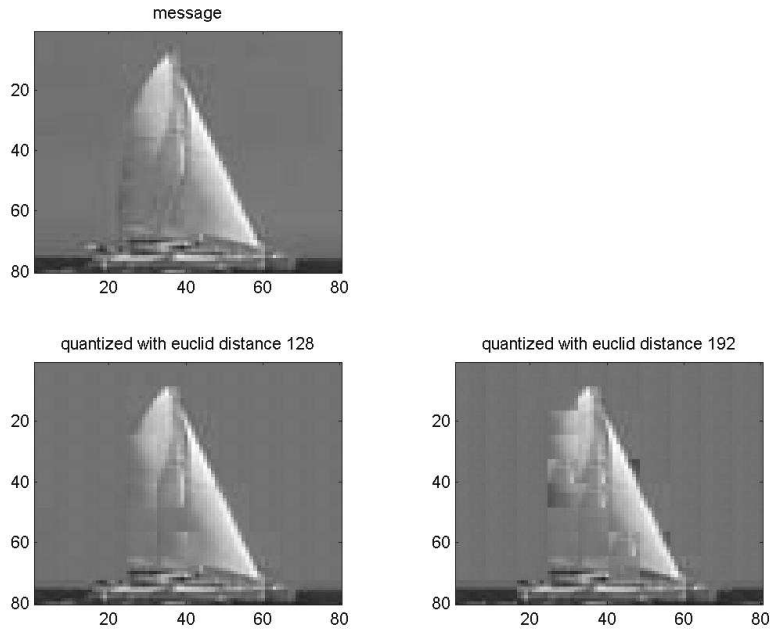

Рис. 2. Визуализация результатов экспериментов.

шее расстояние вносило достаточно сильные искажения в сообщения, а при меньшем расстоянии уменьшались производительность построения книги и объем внедренных данных. Сравнение результатов векторного квантования изображения с различным евклидовым расстоянием представлено на рисунке 2.

Вейвлет- разложение контейнера производилось в базисе Хаара по одному уровню. Многоуровневое преобразование уменьшает количество коэффициентов аппроксимации и объем внедренных данных. Визуализация результатов преобразований изображена на рисунке 3. Максимальный размер части внедренного сообщения без искажений и нарушения прозрачности – 70x70. Для данного изображения достигнуто соотношение объема внедренных данных 1:10. Проводимые эксперименты с другими изображениями- сообщениями показали, что чаще всего это максимальный объем внедренных данных, однако для слабо текстурированных изображений и, следовательно, малого размера кодовой книги объем внедренных данных может доходить до 25%.

При визуализации результатов возникла проблема восстановления цветного изображения. Очертания объектов при восстановлении сохраняются, что достаточно для полутоновой градации яркости. Для цветного же изображения нарушается палитра цветов. Путем экспериментов получен коэффициент $\boldsymbol{\beta}$, на который следует умножить значения полученных отсчетов извлеченного изображения для восстановления первоначальной палитры: $\frac{\boldsymbol{R}}{\boldsymbol{\beta}} = 8$,

где $\boldsymbol{R}$x$\boldsymbol{R}$ – размер внедренной части сообщения.

ЗАКЛЮЧЕНИЕ

В результате проведенных исследований разработан алгоритм встраивания данных с соотношением объема сообщения и контейнера 10% и больше. Встраивание производится в частотной области, что затрудняет атаки на стего. Для разложения контейнера используется вейвлет- преобразование, которое является достаточно производительным и дает возможность получить стего, устойчивое к сжатию в формате JPEG2000. Векторное квантование сообщения позволяет увеличить объем встраиваемых данных, однако снижает производительность встраивания данных.

Предложенный алгоритм может применяться для скрытой передачи информации по общим коммуникационным каналам. Для извлечения необходим контейнер и кодовая книга, которые следует передавать по другим каналам связи или в другое время.

Эксперименты показали, что объем встраиваемых данных зависит в конечном итоге от размера кодовой книги. Развитием этого алгоритма может быть построение адаптивной кодовой книги, которое является достаточно трудоемкой вычислительной задачей.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Chae J.J, Manjunath B.S. A Technique for Image Data Hiding and Reconstruction without Host Image. http://www-iplab.ece.ucsb.edu/publications/99SPIE.pdf.
2. Грибунин В.А. Цифровая стеганография. – М.:Эксмо, 2002.
3. Swanson M.D., Zhu B. Data Hiding for Video- in –Video. Proceeding of IEEE International Conference of Image Processing (ICIP'97), California, Santa Barbara, Oct. 1997, Vol.2, pp.676-679.
4. Mukherjee D., Chae J.J., Mitra S.K. A source and Channel Coding Approach to Data Hiding with Application to Hiding Speech in Video. Proceeding of IEEE ICIP'98, Chicago, Oct. 1998, Vol.1, pp.348-352.
5. Heijmans H., Kamstra L. Reversible Data Embedding Based on the Haar Wavelet Decomposition. Proceeding of VIIth Digital Image Computing: Techniques and Applications, 2003, Sydney.- http://www.cmis.csiro.au/Hugues.Talbot/ dicta2003/cdrom/pdf/0005.pdf.
6. Areepongsa S., Syeld Y.F., Kaewkamnerd N., Rao K.R. Steganography for a Low Bit-Rate Wavelet Based Image Coder. – http://www-ee.uta.edu/dip/paper/icip_2000.pdf.
7. Уэлстид С. Фракталы и вейвлеты для сжатия изображений в действии. –М.: Триумф, 2003.

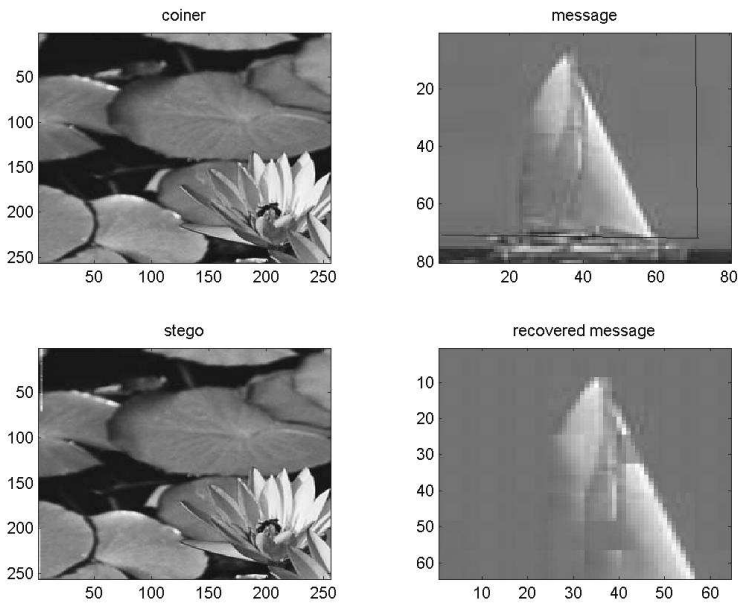

Рис. 3. Внедрение и извлечение изображения.