Эссе по курсу "Защита информации", кафедра радиотехники,
Московский физико-технический институт (ГУ МФТИ),
http://www.re.mipt.ru/infsec

Обзор современных методов сокрытия информации

Дмитриев А.Е., 111 гр.

Стеганография – это наука о передачи секретной информации, причем сам факт передачи остается неизвестен внешнему наблюдателю. Развитие стеганографии мотивируется в основном потребностью защиты интеллектуальной собственности в компьютерных сетях, в основном в интернете. Различают два вида стеганографии:
а) сокрытие информации от пассивного наблюдателя. В этом случае основная цель – не допустить обнаружения скрытой информации.
б) во втором случае информация скрыта от активного противника, т.е. наличие скрытой информации заведомо известно, но получение этой информации противником невозможно. Этот случай распространяется на схемы защиты авторских прав. Здесь используются digital watermarking (цифровые водяные знаки) и fingerprinting (отпечатки пальцев).

В настоящее время методы компьютерной стеганографии развиваются по двум основным направлениям:

1. Методы, основанные на использовании специальных свойств компьютерных форматов;
2. Методы, основанные на избыточности аудио и визуальной информации.

Вкратце о методах первого направления. Здесь применяется метод использования зарезервированных для расширения полей компьютерных форматов данных. Зарезервированные поля имеются во многих мультимедийных форматах, они заполняются нулевой информацией и не учитываются программой. Этот метод очень прост в использовании. Однако явным недостатком этого метода является низкая степень скрытости и передача небольших ограниченных объемов информации.

Для передачи текстовых сообщений используются методы специального форматирования текстовых файлов:

– Методы, основанные на изменении положения строк и расстановки слов в предложении, что обеспечивается вставкой дополнительных пробелов между словами.

– Методы выбора определенных позиций букв (нулевой шифр). Акростих – частный случай этого метода, когда например, начальные буквы каждой строки образуют сообщение или начальные буквы каждого слова.

– Методы, основанные на использовании специальных "невидимых", скрытых полей. Например, использование черного шрифта на черном фоне.

– Методы скрытия в неиспользуемых местах гибких дисков. Информация может записываться, к примеру, на нулевую дорожку.

– Использования имитирующих функций (mimic-function). Метод основан на генерации текстов и является обобщением акростиха. Для тайного сообщения генерируется осмысленный текст, скрывающий само сообщение, расположение букв которого в сгенерированном тексте задается определенным образом.

Все эти методы просты в использовании, но малопроизводительны и обеспечивают низкую степень скрытости. Предназначены только для передачи небольших объемов информации.

Ярким примером применения компьютерной стеганографии является компьютерный вирус Win95.CIH. Этот вирус внедряется в исполняемый файл *.exe. Исполняемый файл может содержать не только код, но и многочисленные дополнительные данные: пиктограммы, различные служебные данные и информация об экспортируемых и

импортируемых функциях. Каждый вид данных, содержащихся в файле, это отдельный объект, занимающий секцию фиксированного размера. Если объект не занимает всего объема секции, то эта часть секции не используется. Поэтому в файле формата PE всегда достаточно свободного места для записи.

Методы использования избыточности аудио и визуальной информации

Младшие разряды представления аудио и видео формата малоинформативны и их изменение практически не сказывается на качестве передаваемого изображения или звука, что дает возможность использования их для кодирования конфиденциальной информации. Но при введении дополнительной информации искажаются статистические характеристики передаваемого файла, что может привести к обнаружению передаваемого сообщения. Для повышения устойчивости к обнаружению применяют методы коррекции статистических характеристик. Основным достоинством данного метода является возможность скрытой передачи большого объема информации, а также возможность защиты авторского права, скрытого изображения товарной марки, регистрационных номеров и т.п.

Наиболее распространенным является метод замены наименее значимых битов или LSB метод. Суть этого подхода заключается в том , что за счет погрешности дискретизации изменение младших разрядов в аудио видео изображении практически не сказывается на качестве передаваемого звука или картинки, особенно если изначально оно было закодировано с большой глубиной передачи цвета. Визуально определить было ли изображение подвергнуто трансформации или нет невозможно, но, используя специальные методы, основанные на статистическом анализе, можно сказать, было ли вкраплено в файл некоторое дополнительное количество информации и даже извлечь ее.

Другие популярные методы встраивания секретных сообщений основаны на использовании форматов файлов с потерей данных (например, JPEG). В отличие от LSB методов они более стойки к геометрическим преобразованиям и обнаружению канала передачи. Это достигается за счет возможности изменять качество сжатых данных в широком диапазоне, что приводит к невозможности определения происхождения изображения.

В современных системах формирования цифровых водяных знаков используется принцип встраивания метки, являющейся узкополосным сигналом, в то время как само маркируемое изображение является широкополосным. Указанный метод реализуется при помощи двух основных алгоритмов и их модификаций. В первом из них используется фазовая модуляция сигнала с псевдослучайной последовательностью чисел для сокрытия секретной информации. Во втором случае весь канал передачи информации делится на несколько каналов и передача осуществляется между ними. На плане исходного изображения встраиваемая метка представляет из себя дополнительный шум со своими статистическими характеристиками. За счет того, что некоторый шум в изображениях присутствует всегда, встраивание метки влияет лишь на уровень имеющегося шума, обычно незаметного для органов чувств. Кроме всего метка является устойчивой к выделению из основного сигнала за счет ее рассеивания по всему частотному диапазону изображения.

Рассмотрим некоторые методы на примере звуковых файлов.

Low-bit coding – это самый простой метод встроить метку в структуру данных. Наименее значимые биты в самплах заменяются на биты встраиваемой метки. Основным недостатком этого метода является его слабая устойчивость к манипуляциям над файлом. В процессе ресамплинга или в результате передачи скрытое сообщение может быть легко искажено или вообще потеряно.

Phase Coding заменяет фазу оригинального звукового сегмента на относительную фазу, которая и представляет собой секретное сообщение. Фаза последовательных сегментов добавляется таким образом, чтобы сохранить относительный фазовый сдвиг между сегментами. Phase coding – один из наиболее эффективных методов сокрытия информации в

терминах отношения уровня сигнала к заметному искажению этого сигнала. Когда отношение фаз между частотами сильно меняется, происходит заметная фазовая дисперсия. Однако “неслышное” кодирование при этом все равно достигается, так как изменение фазы достаточно мало.

Spread spectrum. В обычном канале связи стремятся сконцентрировать информацию в как можно более узком диапазоне частот, чтобы сохранить доступную полосу пропускания и уменьшить энергию, необходимую для передачи. Основной метод расширения спектра спроектирован таким образом, чтобы закодировать поток информации, распространяя секретные данные в как можно большем спектре частот. Это позволяет извлечь скрытую в потоке информацию, даже если происходит интерференция на некоторых частотах. Существует много различных вариантов расширения спектра. Рассмотрим один из них – *Direct Sequence Spread Spectrum encoding* (DSSS). Этот метод располагает скрываемые данные по всему спектру путем умножения его на метку (chip), которая представляет из себя максимальную длину псевдослучайной последовательности, модулируемой на известной частоте. Так как оригинальный сигнал представлен в дискретном формате, то можно использовать sampling rate как частоту метки для кодирования. В результате самая сложная задача в DSSS заключается в установлении правильного начального и конечного квантов для метки, чтобы связывать фазы на соседних звуковых сегментах. Таким образом, возможна генерация метки более высоких частот, а значит получается и более высокая частота ассоциируемых данных. При этом могут быть использованы различные алгоритмы связывания сигнала, но это слишком трудоемкая с вычислительной точки зрения задача.

Echo data hiding. При использовании этого метода данные заключаются в оригинальный звуковой сигнал посредством ввода эха. Данные скрываются при помощи изменения трех параметров эха: начальной амплитудой, степени затухания и задержки. Когда задержка между оригинальным сигналом и эхом уменьшается, сигналы смешиваются. В некоторой точке человеческое ухо уже не может различить эти два сигнала, эхо ощущается как добавочный резонанс. Эту точку достаточно сложно определить, она зависит о качества оригинальной записи, типа записи и слушателя. Обычно слияние сигналов осуществляется в районе 1/1000 секунды, что характерно для большинства типов звуков и слушателей. При кодировании используется две временные задержки, одна для обозначения логической единицы (offset), другая для логического нуля (offset + delta). Обе задержки должны быть меньше порога чувствительности человеческого уха, при котором он может обнаружить эхо. Вдобавок к уменьшению временной задержки необходимо убедиться, что информацию нельзя раскрыть установлением начальной амплитуды и степени затухания ниже порога слышимости человеческого уха.

Для улучшения характеристик сигнала при использовании различных методов сокрытии информации в медиаданных, а также повышения устойчивости полученного сигнала к его анализу и обнаружению скрытой информации применяются некоторые полезные дополнения к обычным алгоритмам стеганографии. Вот некоторые из них.

Adaptive data attenuation. Оптимальный фактор подстройки изменяется при изменении уровня оригинального сигнала. Адаптируя подстройку к небольшим изменениям уровня сигнала или шума, можно удерживать уровень сигнала, представляющего закодированные данные, очень низким в течение интервалов тишины и увеличивать во время интервалов большего уровня звука.

Redundancy and error correction coding. Для того чтобы избежать ошибок при получении сигнала вследствие шума в канале или изменении оригинального звука, полезно применять кодирование с исправлением ошибок к скрываемым данным. Тем не менее, при использовании алгоритмов коррекции ошибок приходиться обходиться компромиссным вариантом, учитывающим надежность данных и объем данных, которые можно при этом скрыть.

Sound context analysis. Обнаруживаемость белого шума, встроенного в оригинальный сигнал, линейно зависит от уровня первоначального звука. Для максимизации объема

скрываемых данных, при условии, что они не будут обнаружены, полезно измерять уровень шума при кодировании. Уровень шума можно определить, измеряя изменение амплитуды сигнала в близлежащих самплах.

Одним из последних методов стеганографии является прием *стохастической модуляции*. На примере графического файла он представляется следующим образом.

Сообщение, которое необходимо передать, обозначим m, состоящее из последовательных 1 и -1(логический 0).

Сначала определяется вероятностная функция $P(x,s) \in \{-1,1\}$, равная 0 только при s=0. Она также должна удовлетворять свойству антисимметричности для всех x: $P(x+s,s)=-P(x-s,s)$. Это свойство полезно в тех случаях, когда значения x+s или x-s выходят за допустимый диапазон значений.

При наложении секретной информации пикселы проходятся в псевдослучайной последовательности, построенной с помощью генератора случайных чисел с распределением, совпадающим с распределением шума, который будет наложен на картинку. Такая последовательность называется стегошумом. При генерации используется специальный стегоключ.

Для каждой точки x генерируется случайной число s. Если s отлично от нуля, то если $P(x+s,s)=m$, то значение пикселя заменяется на x+s, если $P(x+s,s)=-m$, то значение заменяется на x-s. Формально процесс сокрытия есть

$$x'_i=x_i+m_iP(x_i+s_i,s_i)s_i$$

Так как само изображение и стегошум s_i не зависят от секретного сообщения, то сигнал $v_i=m_iP(x_i+s_i,s_i)$ является псевдослучайной последовательностью 1 и -1. Таким образом, v_i имеет такие же статистические свойства, что и стегошум.

Для извлечения закодированного сообщения генерируется стегошум по тому же стегоключу, что и при кодировании. Применяя вероятностную функцию P к пикселам изображения, получаем секретное сообщение, формируемое из ненулевых значений $m_i=P(x_i,s_i)$.

В рассмотренном выше методе используется только один стегошум s_i, который добавляется или вычитается из значений пикселя в соответствии с вероятностной функцией. Возможно получить больший объем скрываемой информации с теми же шумовыми характеристиками. *Улучшенный метод стохастической модуляции* использует сразу два стегошума, добавляя к значениям пикселей изображения всегда либо один стегошум, либо другой, основываясь опять на совпадении бит сообщения и вероятностной функции. Этот метод работает для стегошума с произвольным вероятностным распределением.

Метод стохастической модуляции можно также применять для изображений, полученных с помощью устройств, шум которых зависит от содержания изображения. Подробно этот способ рассмотрен в [2].

Среди последних адаптивных методов стеганографии можно выделить следующие.

Adaptive Least Significant Bit Embedding. Биты секретного сообщения помещаются в наименее значимые биты изображения только в тех пикселях, которые имеют уровень шумовой структуры выше некоторого порогового значения. Уровень шума вычисляется из наиболее значимых бит (MSB) от близлежащих пикселей. Пороговый уровень шума должен быть известен как отправителю, так и получателю для выделения пикселей, несущих информацию.

Statistics-Preserving LSB Embedding. Здесь используются LSB только для пикселей двух цветов c_1 и c_2 с пространственно независимыми статистическими характеристиками. Этот механизм скрытия предназначен для предотвращения атаки, основанной на использовании гистограмм.

Block Parity Embedding. При использовании этого приема все изображение разбивается на небольшие блоки, полностью покрывающие картинку. В каждый блок встраивается не более чем один бит информационного сообщения таким образом, чтобы в блоке выполнялся контроль по четности для всех LSB (xor по всем LS битам). Как и в методе Adaptive Least Significant Bit Embedding для каждого блока вычисляется его сложность (уровень шума по окружающим его блокам) и биты изменяются только в случае, если сложность блока не меньше определенного заранее порога.

Однако, методы LSB даже при их адаптивной модификации неустойчивы к стегоанализу. При передаче сообщений, таким образом, противнику известны места возможного нахождения бит скрытого текста. Следующий метод лишен этих недостатков, но при этом он более сложен в реализации и использовании.

Perturbed quantization. Оригинальное изображение сначала подвергается обработке с потерей данных (к примеру, квантование при JPEG компрессии). Этот процесс обычно представляет последовательные операции преобразования с плавающей точкой и квантования. Наибольшие погрешности квантования происходят для значений близких к середине интервала квантования. Из-за шума, который обычно присутствует в цифровых изображениях, погрешность квантования этих значений подавляется шумом, что полностью повторяет случайный процесс. Отправитель скрытого сообщения немного меняет параметры процесса квантования, чтобы скрыть биты сообщения. Так как процесс преобразования происходит с потерей информации, то аналитик не может легко восстановить те мелкие детали изображения, которые позволили бы ему собрать статистику для элементов, которые были “неправильно” квантованы.

Процесс PQ может быть осуществлен несколькими способами. Во-первых, при помощи операции, сопровождающейся потерей информации. К таким относятся resizing, уменьшение глубины цвета на некоторое количество бит, JPEG сжатие. Во-вторых, использование несколько измененных коэффициентов квантования, в зависимости от того, насколько близко лежат заданные коэффициенты к середине интервала квантования. Также известен способ встраивания скрытой информации в jpeg файл при его рекомпрессии с ухудшением качества.

На сайте SecurityLab можно найти разнообразные программы, предоставляющие функциональные возможности современных методов стеганографии.

http://www.securitylab.ru/tools/22202.html

Ссылки:

[1] Techniques for data hiding, by W. Bender, D. Gruhl, N. Morimoto, A. Lu
IBM SYSTEMS JOURNAL, VOL 35, NOS 3&4, 1996

[2] Digital image steganography using stochastic modulation, Jessica Fridrich and Miroslav Goljan, 2003
http://www.ws.binghamton.edu/fridrich/Research/stochastic_modulation02.pdf

[3] Writing on Wet Paper, Jessica Fridrich, Miroslav Goljan, Petr Lisonek, and David Soukal, 2005
http://www.ws.binghamton.edu/fridrich/Research/EI5681-33_WPC.pdf

[4] КОМПЬЮТЕРНАЯ СТЕГАНОГРАФИЯ ВЧЕРА, СЕГОДНЯ, ЗАВТРА, Барсуков В.С., Романцов А.П., 2000
http://st.ess.ru/publications/articles/steganos/steganos.htm