Стеганография в спаме

Стеганография.

Буквально в переводе с греческого стеганография означает «тайнопись»(steganos - секрет, тайна; graphy - запись). Как и криптография, стеганография служит для секретной передачи сообщений. Но в отличие от криптографии, которая основывается на методах, способных исказить исходное сообщение настолько, что криптоаналитик не сможет его реконструировать, методы стеганографии позволяют скрыть сам факт существования сообщения.

Изобретение «тайнописи» приписывают древним грекам. Известны два хорошо документированных метода, которые ими применялись. Первый метод заключается в том, чтобы написать скрываемое на бритой голове раба или слуги и посылать, только тогда, когда отрастут волосы. Второй основывался на деревянных табличках, покрытых слоем воска, которые использовались для письма. Сообщение вырезалось непосредственно по дереву, а затем оно покрывалось слоем воска, что создавало видимость отсутствия всякого сообщения.

Если попытаться определить то, какие требования предъявляются к стеганограмме, то можно выделить следующее:

- Передача носителя стеганограммы является обыденным делом;
- Добавления сообщения в носитель не должно быть заметно без знания ключа или метода, при помощи которого это сообщение было встроено.

Из последнего требования можно выделить следствие, что объём встраиваемого сообщения должен быть много меньше объема носителя этого сообщения.

Стеганограммы в текстовых данных

Стеганограммы в текстовых данных основываются на факте избыточности языка. Продемонстрировать это можно множеством способов. Остановимся лишь на тех методах что были использованы в стеганографии.

Смещение в словах

Если взять вполне безобидную фразу:

```
Apparently neutral's protest is thoroughly discounted and ignored. Isman
hard hit. Blockade issue affects pretext for embargo on by-products, ejecting
suets and vegetable oils.
```

А затем выписать вторые букву каждого слова:

```
Apparently neutral's protest is thoroughly discounted and ignored. Isman
hard hit. Blockade issue affects pretext for embargo on by-products, ejecting
suets and vegetable oils.
```

То полученное сообщение носит далеко не простой смысл:

```
Pershing sails from NY June I
```

Это иллюстрирует метод использованный немецким шпионом во время Второй Мировой Войны. Явным преимуществом данного метода является то, что в среднем на одну букву исходного текста приходится всего 6 букв стеганограммы, но недостатки также очевидны. Составление подобного сообщения требует значительных усилий стеганографа по составлению связанной конструкции из слов, у которых определённые буквы должны иметь заданное значение.

Нули и единицы

Если перед нами стоит задача спрятать произвольную битовую последовательность, то текстовые данные представляют множество способов как это сделать.

Самым простым способом является использование неоднозначности ASCII-кодов. Символ пробел, закодированный кодом 32 и кодом 255 неотличим в тексте, что может быть использовано для передачи одного бита информации. Ещё один бит информации можно включить при помощи перехода на новую строку, который может быть записан различными символами при помощи кодов 12 и 15, неотличимыми при отображении на мониторе.

Если первый недостаток этого метода лежит в области удобности использования и заключается в низком соотношении размера передаваемого сообщения к размеру носителя, для такого способа оно составляет 1/50, то другие недостатки являются характерно-стеганографичным. Во-первых, нет никаких объективных причин для того, чтобы кодировать одни и те же символы разными ASCII кодами в одном и том же исходном тексте, а значит, будучи замеченным, это сразу вызовет подозрение. Во-вторых, если подобная стеганограмма на пути следования попадет в программу обработки текста, то велика вероятность, что она будет искажена, так как незнакомые символы могут быть исключены или заменены на их разрешенные аналоги. В итоге перед получателем может предстать уже неумышленно очищенный от исходного сообщения носитель.

Более незаметные и эффективные методы основываются на избыточности языка, как такового. Нули и единицы можно прятать в синонимах слов. Синонимы, находящиеся в словаре раньше, могут означать ноль, а слова, находящиеся в конце словаря, могут означать единицу. Также бит данных можно различить в «птица, сидевшая на ветке» и «сидевшая на ветке птица». Порядок слов, количество слов в предложениях – язык предоставляет множество возможностей для стеганографа, но их детальное рассмотрение лежит за рамками данной работы.

Спам

Спамом называют нежелательную корреспонденцию рекламного характера. Также рассылка спама носит массовый характер, то есть одно и то же сообщение получают тысячи человек. Хотя особенно эффективной подобную рекламу не назовёшь, её копеечная стоимость делает её всё более и более популярной. За экспоненциальным ростом количества спама рост количества обыкновенных сообщений явно не успевает, что привело к тому, что в прошлом году по разным оценкам от трети до половины электронных сообщений были спамом. Спам становится всё более и более **обычным**, что несомненно делает 24 миллиона сообщений в день прекрасным носителем для новых методов стеганографии. За счет ещё большей избыточности языка при помощи лексической реконструкции можно достигнуть прекрасных результатов, как это было сделано в **Spammimic**(**www.spammimic.com**). Но кроме этого сама нежелательная природа спама приводит к постоянным попыткам сократить количество спама, достигающее конечного пользователя. Это привело к тому, что спаммеры стали оборачивать само сообщение в защитный камуфляж, призванный надежно оградить сообщение от всевозможных ловушек. С такой точки зрения можно сказать, про стеганографию и спам, что в обоих есть носитель, и есть сообщение, которое встроено в этот носитель, но если задача стеганографа, чтобы это сообщение попало только выделенным получателям, минуя проверки стегоаналитиков, то задачей спаммера является как раз получение сообщения как можно большим числом людей, при этом на носитель также возлагается задача скрытия природы сообщения от всевозможных анти-спам фильтров. Очевидно, что методы стеганографии, использующие спам, как носитель, должны быть ориентированы, как раз, на встраивание сообщений в камуфляж спам-сообщения. Ещё стоит отметь, что сам метод рассылки спама надежно разделяет отправителя и получателя. Если выяснить кто отправитель ещё возможно, при неосторожности с его стороны, то узнать кто из адресатов настоящий получатель информации – задача нетривиальная.

Spammimic(www.spammimic.com)

Если взять почти любой образец рекламного сообщения, то его структуру можно представить в виде набора отдельных блоков:

- Приветствие;
- Объявление причин, по мнению отправителя, достаточных для того, чтобы разослать это сообщение;
- Коммерческое предложение;
- Явный выигрыш который для получателя сообщения;
- «Опыт использования» коммерческого предложения другими «получателями».

Конечно что-то может исчезнуть, может появится что-то новое, но основа остается практически всегда неизменной. Но главное преимущество в том, что эти блоки могут быть практически независимы друг от друга по смыслу. Написав произвольные сообщения для каждого из блоков, выдержав их в общем стиле, можно затем составлять рекламные сообщения из любой комбинации исходных блоков.

Подобные рассуждения привели к созданию Spammimic. Последовательность из 100 бит превращается в 13 предложений рекламного характера. В итоге мы получаем стеганограмму, размер которой приблизительно в 50 раз превышает размер исходного сообщения. Так, например, слово **steganograph** превращается в следующее сообщение:

```
Dear Business person , This letter was specially selected
to be sent to you . We will comply with all removal
requests . This mail is being sent in compliance with
Senate bill 1916 , Title 1 , Section 305 . THIS IS
NOT MULTI-LEVEL MARKETING . Why work for somebody else
when you can become rich as few as 59 days ! Have you
ever noticed more people than ever are surfing the
web plus most everyone has a cellphone ! Well, now
is your chance to capitalize on this . We will help
you turn your business into an E-BUSINESS and decrease
perceived waiting time by 130% ! You can begin at absolutely
no cost to you ! But don't believe us . Ms Simpson
of Hawaii tried us and says "I was skeptical but it
worked for me" ! We are a BBB member in good standing
. Do not delay - order today . Sign up a friend and
your friend will be rich too ! Best regards !
```

Сообщение несколько бессмысленно? Это обычно для спама.

К чему приводит борьба со спамом

Как было сказано выше, количество спама постоянно растет. Развиваются и методы, призванные сократить количество спама, которое доходит до конкретного электронного почтового ящика. Спам, отправленный без подобающего камуфляжа, будет почти наверняка отбракован любым современным средством по фильтрации спама. Такая судьба, например, ждёт сообщения, создаваемые при помощи Spammimic. Как сделать спам незаметным для фильтров? Как правильно нанести камуфляж? Рассмотрим те ответы, которые всё же приходят к нам в почтовые ящики, несмотря на множество изощрённых барьеров.

Замена букв

Не так то просто отличить русскую букву «о» от его английского аналога «o». Если выписать все буквы русского алфавита, имеющие похожие по начертанию буквы в английском, то мы получим:

А В Е К Н О Р С Х

a e o p c x y

В итоге фильтры, которые ловили сообщение со словом «гербалайф», будут чаще ошибаться, так как это слово можно написать шестнадцатью разными способами, в разных комбинациях заменяя буквы «г**ерб**а**л**а**й**ф». Но что это дает для стеганографа? Если посчитать частоту, с которой эти символы встречаются в русском языке, то получается, что заменяемыми являются 33,9% букв в произвольном тексте, а следовательно можно каждую третью букву текста использовать для кодирования битовых последовательностей. Если принять, что для кодирования одной буквы в битовых последовательностях нужно 5 бит($2^5 = 32$ буквы – русский алфавит, в котором буквы «**е**» и «**ё**» считаются за одну букву), то получается, что на 15 букв исходного текста можно наложить одну букву секретного послания.
Рассмотрим стеганограмму:

```
Уверен, что если ты читаешь дальше, то ты не из тех, кто пытается выдрать заветный кусок хлеба…
```

А теперь выделим жирным шрифтом заменённые буквы, а курсивом не замененные:

Ув**е**рен, что **ес**ли ты чит**а**ешь д**а**льше, *то* ты *не* из те**х**, кт**о** пыт**а***ется* выд**р**ать з**а**ветный к**у***сок хлеба*…

В итоге битовая последовательность, скрытая в сообщении составляет:

```
10011 10100 00111 00101 01000 00
```

Если воспользоваться кодированием, когда буква обозначается своим порядком в алфавите, то мы получим слово «**стего**».

Микроточки XXI века
Как видно, развитие спама можно описать словами: «Оставаясь почти неизменным визуально, он должен принимать каждый раз новую, непохожую на что-либо прежде, форму». Небольшим отклонением от этого стал следующий метод спаммеров. Основная его идея заключается в том, что если некоторые буквы повторять дважды, то старые слова принимают новый вид, но при этом могут быть прочитаны без затруднений. Но двойные буквы не должны идти слишком часто, иначе воспринимать текст будет труднее. Это ощутимый недостаток как для спаммеров, так и для стеганографов, которые решат воспользоваться этим методом. Вызовет подозрение и использование для кодирования символов с определенным, пусть и достаточно большим, периодом. В таком случае стеганографу можно предложить выход в использовании кодов с значительно высокой вероятностью следования нуля, чем единицы(если удвоенная буква означает единицу). Простейший способ это сделать – это взять три бита и обозначить 000 за 10000000 и так далее вплоть до последовательности 111, которую обозначить 00000001. В итоге мы получаем стеганограмму, которую не отличишь от обычного спама. При таком подходе содержание сообщения в носителе составляет приблизительно 1 к 14.
Но этот метод слишком слабо защищал спам от того, чтобы быть незамеченным. Тогда на помощь пришёл язык разметки гипертекста (HTML). Вообще HTML может служить инструментом стеганографа сам по себе. Множество атрибутов тегов, которые автор может заполнять по желанию, закрывающие теги, которые иногда можно опускать, - всё это потенциальный носитель для секретных сообщений. Но не это привлекло спаммеров в HTML. Язык разметки даёт потрясающие возможности иметь в HTML вполне приличных размеров текст, но не показать его пользователю. Рассмотрим пример:

```
Сте<span style='display:none'>га</span>нограмма
```

Что перед нами? С точки зрения текстовой информации – это стеганограмма, но если отобразить этот фрагмент, как HTML, то это всего лишь безобидная стенограмма. Это

позволяет спаммеру разделить рекламное сообщение и средства для его камуфлирования от фильтров. Стеганографу же это развязывает руки в плане применения методов для встраивания информации в носитель. Антифильтровый камуфляж не может полностью состоять из передаваемого сообщения, но в таком случае задача встраивания сообщения может решаться уже методами кодирования, а не стеганографии.

Спам в картинках

Одним из самых последних способов камуфлирования спама, который набирает всё большую популярность является передача рекламного сообщения в виде изображения. Этот метод ставит в тупик все самые современные фильтры. Создавать такие сообщения в разы проще, чем автоматизировать их анализ. При этом встраивание сообщения в подобный камуфляж является классической задачей компьютерной стеганографии. Правда стеганоанализ цифровых изображений – тоже задача классическая, но всё возрастающее количество спама позволяет говорить о том, что сообщение скрытое в изображении-спаме на порядки незаметнее, чем изображение-стеганограмма, посланная непосредственно адресату.

Методы стеганографии в цифровых изображениях основываются на том, что малые изменения в цвете отдельных пикселей изображения, при использовании форматов BMP или GIF, или изменения в незначительных коэффициентах разложения, как в формате JPEG, незаметны для человеческого глаза. Существует множество программных средств, которые служат для скрытия информации в цифровых изображениях. Но отсутствие стандартов на стеганографические алгоритмы, динамичное изменение алгоритмов стеганографии, призванное оградить от динамичного изменения алгоритмов стеганоанализа говорит о научной молодости молодости компьютерной стеганографии по отношению к компьютерной криптографии.

Литература

*О. В. Генне «*ОСНОВНЫЕ ПОЛОЖЕНИЯ СТЕГАНОГРАФИИ» *журнал "Защита информации. Конфидент", №3, 2000*

Anderson, Ross and Fabien Petitcolas. “On the Limits of Steganography.” *IEEE Journal of Selected Areas in Communication,* 16(4):474-481, May 1998.