УДК 621.391

Б.Я. Рябко, А.Н. Фионов

Идеальные стенографические системы

Предлагаются две базовые конструкции стеганографических систем, гарантирующие невозможность выявления факта наличия в стегоконтейнере скрытой информации.

Основная (классическая) задача стеганографии — передать сообщение так, чтобы сам факт передачи оставался скрытым. Как правило, задача скрытия факта передачи сообщения решается путем встраивания сообщения в некоторый «контейнер», например графический файл, передача которого осуществляется на регулярной основе и не вызывает подозрений. Исходный контейнер называется пустым, а контейнер с внедренным сообщением — заполненным (в реальных стегосистемах контейнеры могут заполняться лишь на некоторую долю от их емкости). Кроме того, в современных системах сообщение обычно шифруется и встраивается в контейнер в зашифрованном виде. Обратная задача, т.е. задача выявления наличия скрытого сообщения, составляет предмет стегоанализа. В основе стегоанализа лежит тот факт, что внедренное сообщение нарушает «естественные» информационные связи и зависимости в контейнере, в частности, искажает его статистическую структуру. В [1] было сформулировано понятие идеальной стегосистемы, не позволяющей определить факт наличия скрытого сообщения. Это понятие близко к соответствующим понятиям теории идеальных шифров К. Шеннона и базируется на теоретико-информационных принципах. Для того чтобы построить идеальную стегосистему, необходимо разработать такие методы внедрения сообщений, которые не искажали бы статистическую структуру контейнера. Иными словами, пустой и заполненный контейнеры должны быть статистически неразличимы. В данной статье представлены новые эффективные методы решения этой задачи. Заметим, что стеганография и стегоанализ развиваются параллельно, причем достижения в одной области обогащают другую. Так, развиваемые в данной работе подходы к построению идеальных (невскрываемых) стегосистем демонстрируют, с одной стороны, границы применимости методов стегоанализа, а с другой — показывают, как проводить стегоанализ в практической ситуации, когда используемый контейнер отличается от той модели, для которой была построена идеальная стегосистема.

Первая предлагаемая конструкция идеальной стегосистемы была представлена в [2] и базируется на нумерации множества равновероятных последовательностей. Обозначим последовательность элементов контейнера, в которые непосредственно будет внедряться информация, через $x = x_1x_2x_3\ldots$, а внедряемое зашифрованное сообщение — через $y = y_1y_2y_3\ldots$. Разобьем последовательность x на блоки некоторой фиксированной длины, обозначим один такой блок через u и рассмотрим множество S_u равновероятных последовательностей, одним из элементов которого является u. Например, если $u = acb$ и известно, что буквы независимы и одинаково распределены, то $S_u = \{abc, acb, bac, bca, cab, cba\}$. Основная идея состоит в использовании битов зашифрованного сообщения в качестве номера, по которому из S_u выбирается последовательность v, которая затем записывается в контейнер вместо u. Так как последовательности v и u равновероятны, то статистическая структура контейнера не изменяется, что дает в результате идеальную стегосистему. Небольшая техническая проблема состоит в том, что с помощью битов зашифрованного сообщения можно задавать номера, равномерно распределенные в множествах, мощность которых равна степени двух, в то время как мощность S_u может быть не равна степени двух. Предлагаемое решение состоит в разбиении S_u на подмножества, мощность которых равна степени двух, и выбора одного такого подмножества с вероятностью, пропорциональной его мощности. Для рассмотренного примера S_u разбивается на два подмножества $\{abc, acb, bac, bca\}$ и $\{cab, cba\}$, первое выбирается с вероятностью $4/6$, второе — с вероятностью $2/6$. Если выбрано первое подмножество, то для выбора последовательности используются два бита сообщения, т.е. два бита скрываются в текущем блоке контейнера. Если выбрано второе подмножество, то используется только один бит сообщения.

Второй подход к построению идеальных стегосистем основывается на использовании непосредственного вероятностного описания (обычно в виде условных вероятностей) последовательности элементов контейнера. Данный подход может применяться, когда трудно построить множество равновероятных последовательностей и провести его нумерацию. Идея заключается в кодировании зашифрованного сообщения кодом, в котором алфавит и вероятности появления кодовых символов полностью соответствуют алфавиту и вероятностям символов последовательности x контейнера. Для решения этой задачи могут быть применены так называемые коды со стоимостью (например, арифметическое декодирование), но они не позволяют получить точно требуемой вероятности появления ко-

довых символов. В [3] предложены коды, которые за счет рандомизации обеспечивают точно заданное распределение кодовых символов. Закодированное таким кодом сообщение просто записывается в контейнер на место последовательности x. Для иллюстрации идеи построения кода рассмотрим случай двоичного алфавита контейнера $A = \{a, b\}$, и пусть $p = p(a) > p(b)$. Буквам источника (0, 1) и кодовым буквам $(a,\ b)$ выделяются интервалы, равные их вероятностям: $I(0) = [0;\ 0,5)$, $I(1) = [0,5;\ 1)$, $I(a) = [0,\ p)$, $I(b) = [p,\ 1)$. На первом шаге, если входной бит $y_1 = 0$, мы выдаем на выход a (так как $I(0) \subset I(a)$), но если $y_1 = 1$, то выдаем на выход a или b с вероятностями, пропорциональными их долям в $I(1)$. Выдача b завершает алгоритм (так как $I(b) \subseteq I(1)$). В противном случае вычисляем новое распределение внутри предыдущего $I(a)$ (умножаем вероятности на вероятность символа a) и получаем $I(a) = [0,\ p^2)$, $I(b) = [p^2,\ p)$. Далее принимаем решение, какой кодовый символ выдать на выход, точно так же, как и на первом шаге. Число шагов в алгоритме не ограничено, но их среднее число не велико. В результате каждый бит сообщения кодируется несколькими различными кодовыми словами, распределение кодовых символов может меняться на каждом шаге, но гарантируется точное соответствие вероятностей их появления заданным.

Литература

1. Cachin C. An information-theoretic model of steganography // Lecture Notes in Computer Science (Proc. 2nd Information Hiding Workshop). – Springer Verlag, 1998. – Vol. 1525. – P. 306–318.

2. Ryabko B., Ryabko D. Information-theoretic approach to steganographic systems // IEEE International Symposium on Information Theory. Nice, France, 2007. – P. 2461–2464.

3. Fionov A., Ryabko B. Simple ideal steganographic system for containers with known statistics // XI International Symposium on Problems of Redundancy in Information and Control Systems. – St.-Peterburg, 2007. July 2–6. – P. 184–188.

Рябко Борис Яковлевич
ГОУ ВПО Сибирский государственный университет телекоммуникаций и информатики
Проректор по научной работе, д.т.н., профессор
Эл. почта: boris@ryabko.net.

Фионов Андрей Николаевич
ГОУ ВПО Сибирский государственный университет телекоммуникаций и информатики
Д.т.н., профессор кафедры прикладной математики и кибернетики
Эл. почта: fionov@neic.nsk.su.

B.Ya. Ryabko, A.N. Fionov
Ideal stegographic systems.

Two base designs stenographic the systems, revealing of the fact of presence guaranteeing impossibility in stegocontaner are offered the latent information.
