УДК 519.24

Простое построение совершенных стегосистем на основе различных ошибок в помехоустойчивых кодах в модели трех каналов

© П.В. Слипенчук

МГТУ им. Н.Э. Баумана, Москва, 105005, Россия

Дано описание модели трех каналов, применимой для стегосистем в помехоустойчивых кодах. Объяснено, что модель трех каналов является частным случаем стеганографической модели Кристиана Кашена. Даны определения совершенной и идеальной стегосистемы. Доказано, что любая идеальная стегосистема при соблюдении двух очевидных условий является совершенной (от пассивного противника). В качестве примера совершенной стегосистемы приведена математическая модель возникновения ошибок на оптических дисках. На основе этой модели разработан алгоритм стеганографии в кодах исправления ошибок, использование которого делает стегосистему совершенной. Приведен формальный алгоритм вкрапления стегосообщения.

Ключевые слова: *стеганографические модели, стеганография, коды исправления ошибок, совершенная стегосистема, идеальная стегосистема, лемма о совершенности идеальной стегосистемы.*

Введение. Еще в 1998 г. Кристиан Кашен ввел понятие *совершенной стегосистемы* (от пассивного противника) (perfectly secure stegosystem against passive adversaries) [1]. С тех пор неоднократно предпринимались попытки построить совершенную стегосистему для различных типов стеганографических приемов.

Для модели трех каналов [2], если в канале C_2 меньше помех, чем в канале C_1, можно легко построить идеальную стегосистему (далее приводится пример).

В настоящей статье будет доказано, что при простых допущениях *любая идеальная стегосистема* [2] *является совершенной*. Это значит, что для построения совершенной стегосистемы достаточно доказать ее идеальность. Таким образом, для класса стеганографии в помехоустойчивых кодах построение совершенных стегосистем возможно с помощью построения идеальной стегосистемы в модели трех каналов.

Указанный подход применим и для любых других каналов передачи данных или носителей информации, использующих помехоустойчивые коды, например для таких, как Wi-Fi, спутниковая связь, твердотельный накопитель (Solid-State Drive, SSD), накопитель на

жестких магнитных дисках (НЖМД, hard disk drive, HDD), флеш-накопители и т. п.

Модель стеганографической системы Кашена и совершенная стегосистема. Кристиан Кашен предлагает стеганографическую модель под названием «модель стегосистемы с секретным ключом» (The model of a secret-key stegosystem) [1]. На рис. 1 показан принцип работы этой стегосистемы.

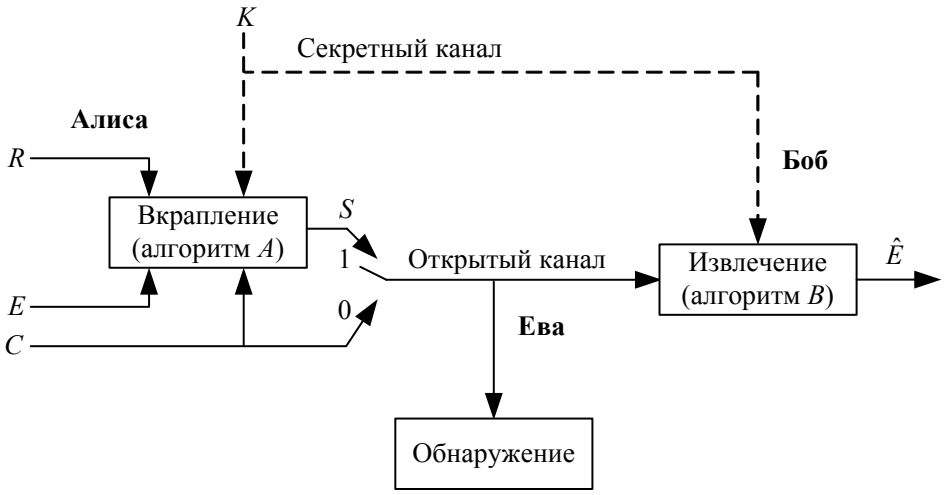

Рис. 1. Модель стегосистемы с секретным ключом – стеганографическая модель Кристиана Кашена

Переключатель S определяет состояние Алисы:

• *пассивное состояние* (переключатель в позиции 0) — Алиса отправляет только пустые контейнеры C Бобу по открытому каналу передачи данных. Ева имеет возможность просматривать пустые контейнеры C;

• *активное состояние* (переключатель в позиции 1) — Алиса отправляет стегосообщение E, которое она вкрапила в пустой контейнер C, используя алгоритм A. Имея на входе контейнер C (получаемый случайным источником пустых контейнеров R), ключ K и сообщение E, алгоритм A выдает на выходе стегоконтейнер S. Стегоконтейнер отправляется по открытому каналу Бобу. Противник Ева и получатель Боб принимают S. Используя алгоритм извлечения B, Боб с помощью ключа K извлекает сообщение $\hat{E} \in \boldsymbol{E}$ из S в надежде, что он получил искомое стегосообщение от Алисы $\left(\hat{E} = E\right)$.

У Боба имеется так называемый «оракул»[1], с помощью которого он определяет, активна Алиса или пассивна. Если Боб знает, что Алиса активна, то он извлекает стегосообщение, в противном случае он не использует алгоритм извлечения B.

Задача Евы определить, когда Алиса передавала пустой контейнер, а когда стегоконтейнер.

Обозначим через M сообщение в канале. Если Алиса активна, то $M = S$ (стегоконтейнер), а если пассивна, то $M = C$ (пустой контейнер). Определим распределения $P_C(y)$ и $P_S(y)$ соответственно как вероятность появления пустого контейнера $y \in \boldsymbol{C}$, если передавался пустой контейнер, и вероятность появления стегоконтейнера $y \in \boldsymbol{S}$, если передавался стегоконтейнер.

Допустим, что P_C и P_S известны Еве. Это напоминает предположение о неограниченных вычислительных ресурсах, которыми обладает противник, при определении Шенноном совершенной криптосистемы в работе [3]:

$$D\left(P_{X1}\middle\|P_{X2}\right) = \sum_{x\in\boldsymbol{X}} P_{X1}(x)\log_2\frac{P_{X1}(x)}{P_{X2}(x)}. \qquad (1)$$

Введем множество $\left\{\hat{\mathbb{R}} = \mathbb{R}\cup\{\infty\}\right\}$. Величину $D\left(P_{X1}\middle\|P_{X2}\right)$, заданную формулой (1) из множества $\boldsymbol{X}$ в множество $\hat{\mathbb{R}}$, называют *относительной энтропией* (relative entropy), если определить $0\log_2\frac{0}{0} = 0$; $T\log_2\frac{T}{0} = \infty$. Если хотя бы одно слагаемое в формуле (1) равно ∞, то $D\left(P_{X1}\middle\|P_{X2}\right) = \infty$. Заметим, что относительная энтропия не является симметричной величиной, иначе говоря, $D\left(P_C\middle\|P_S\right) \neq D\left(P_S\middle\|P_C\right)$.

Система называется *совершенной* (*от пассивного противника*), если относительная энтропия между P_C и P_S равна нулю, т. е. если

$$D\left(P_C\middle\|P_S\right) = 0. \qquad (2)$$

Система называется ε-*секретной* (*от пассивного противника*), если

$$\left|D\left(P_C\middle\|P_S\right)\right| \le \varepsilon. \qquad (3)$$

[1] Термин «оракул» использует сам Кашен: «К тому же мы предполагаем, что у Боба имеется оракул, который говорит ему, активна Алиса или нет. Это серьезное допущение, обозначим его как одно из основных свойств безопасности стегосистемы. Не принимая данное допущение, мы не ухудшим качество системы. Действительно, если Боб попытается извлечь сообщение из контейнера, когда Алиса пассивна, он получит всего лишь «мусор» [1, с. 31] (*перевод автора*).

Если система совершенная, это означает, что, получая на вход контейнер, мы не можем с вероятностью, отличной от 0,5, определить, принадлежит ли он к стегоконтейнеру или к пустому контейнеру.

Модель трех каналов и идеальная стегосистема. Описание модели трех каналов впервые приведено в работе [2]. В данной модели выступают пять сторон: Алиса, Боб, Алена, Борис и Иванов. Одна из задач Иванова определить, какая пара использует стеганографию при передаче сообщения: Алиса и Боб или Алена и Борис.

Алена и Борис передают друг другу контейнеры, не содержащие стегосообщения (пустые контейнеры) (рис. 2). Для этого Алена использует *помехоустойчивый код*. Перед отправкой *информационный вектор* подается на кодер A, который выдает на выходе *кодовый вектор*. После прохождения кодового вектора через каналы C_1 и C_3 Борис декодирует его и получает *информационный вектор*.

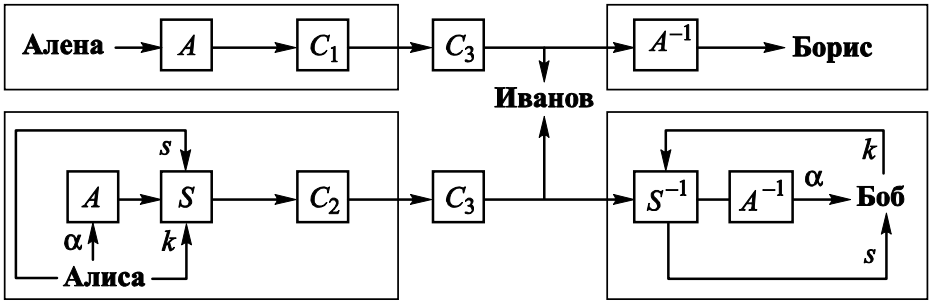

Рис. 2. Модель трех каналов

Алиса перед отправкой кодового вектора вкрапляет в него сообщение s. Таким образом, *кодовый вектор* (получаемый из информационного вектора α) выступает в качестве *контейнера для стегосообщения,* а сообщение представляет собой *искусственные ошибки* [2]. Позиции, содержащие стегосообщение, определяются неким алгоритмом S, зависящим от ключа k. Таким образом, Боб, зная ключ k и алгоритм S, может отличить искусственные ошибки от подлинных, а Иванов не может.

Заметим, что модель трех каналов есть частный случай модели Кашена. Действительно, в модели трех каналов Алиса — это то же самое, что Алиса в активном состоянии в модели Кашена, а Алена — это Алиса в пассивном состоянии. Задача Иванова — определить, кто передает скрытые сообщения — Алиса или Алена (см. рис. 2); задача Евы — определить, когда Алиса активна (см. рис. 1). Пусть на вход канала C_i подается последовательность битов длиной n (см. рис. 2). Каждую ошибку можно представить n-ричным числом в двоичной системе счисления. Для каждой последовательности битов введем вероят-

ность возникновения данной ошибки. Сумма всех ошибок (с учетом нулевой ошибки) равна единице.

Будем обозначать распределения ошибок длиной n, которые представляют собой упорядоченные вероятности возникновения ошибок в каналах C_1 и C_2, как R_{C_1} и R_{C_2}. Ошибка является случайной величиной. Обозначим через e_i ошибку в канале C_i.

Рассмотрим *искусственные ошибки*, вкрапляемые в *кодовый вектор* перед каналом C_2. Ошибку будем обозначать буквой v. После прохождения стегоконтейнера через канал C_2 на выходе будем иметь стегоконтейнер, содержащий и искусственные, и подлинные ошибки. Ошибка, состоящая из искусственных и подлинных ошибок, представляет собой случайную величину $e_2 + v$. Определим $x + y$ как побитную сумму по модулю 2. Множество всех ошибок длиной n будем обозначать $\boldsymbol{X}$. Для $e_2 + v$ справедлива формула

$$P\left(e_2 + v = x\right) = \sum_{\forall y \in \boldsymbol{X}} P\left(e_2 = x + y\right) \cdot P\left(v = y\right). \qquad (4)$$

Обозначим через R_1 распределение случайной величины e_1, а через R_2 распределение случайной величины $e_2 + v$. Таким образом, R_1 — это распределение *пустых контейнеров*, проходящих через канал C_1, а R_2 — распределение *стегоконтейнеров*, проходящих через канал C_2.

Стегосистему S назовем *идеальной стегосистемой* (*в кодах, исправляющих ошибки) для канала* C_2 *по отношению к каналу* C_1, если распределения R_1 и R_2 совпадут [2].

Совершенность идеальной стегосистемы. Введем ряд обозначений. Канал C_1 будем обозначать буквой C, а канал C_2 — буквой S. Соответственно y_C и y_S — это *информационные векторы*, которые подаются на вход *кодера исправления ошибок* в модели трех каналов для каналов C и S. Обозначим через $Y_Z\left(y\right)$ вероятность выбора информационной матрицы y для некоего канала Z.

Предположим, что пустые контейнеры в канале C всегда декодируются верно. Действительно, в случае канала C (не стеганографический канал), если кодовая матрица с существенной вероятностью декодируется неверно, создатели помехоустойчивого кода разработали «плохой» код. Данный код не имеет практического смысла. Мы считаем, что помехоустойчивый код декодирует искомую информационную матрицу с почти достоверной вероятностью. Данное предпо-

ложение автор считает разумным, тем не менее формально оно является условием леммы.

Предположим также, что $Y_C(y) = Y_S(y)$. Смысл этого предположения в следующем: мы считаем, что Алиса и Боб хорошо изучили пользователей канала Алену и Бориса. По этой причине «они не выделяются из толпы» и используют в качестве информационной матрицы (из которой они делают контейнер для вкрапления скрытых сообщений) такие же информационные матрицы, что и Алена с Борисом. Как и первое условие леммы, данное предположение будем считать очевидным допущением.

Лемма. *Предположим, что пустые контейнеры в канале C всегда декодируются верно и* $Y_C(y) = Y_S(y)$. *Тогда любая идеальная стегосистема является совершенной.*

◄ *Доказательство.* Распределения R_1 и R_2 переобозначим как R_C и R_S соответственно.

Для доказательства того, что идеальная стегосистема является совершенной, нам необходимо проверить справедливость формулы (2).

Найдем вероятность появления некоторого *пустого контейнера* x_C и некоторого *стегоконтейнера* $x_S = x_C$. Множество всех пустых и стегоконтейнеров обозначим $\boldsymbol{X}$, а векторы ошибок, произошедших в каналах с шумом *S* и *C*, на выходе которых получили *кодовые матрицы* x_S и x_C (см. [2]), — соответственно e_S и e_C. Так как $x_S = x_C$ и по условию леммы кодовая матрица декодируется верно в информационную матрицу, то равенство $e_S = e_C$ справедливо.

Обозначим $E_Z(e)$ вероятность появления ошибки *e* в канале *Z*. Так как система по определению идеальная, то $R_C = R_S$. Из этого следует, что

$$E_C(e) = E_S(e). \tag{5}$$

Так как $e_S = e_C$ и $x_S = x_C$, то $y_S = y_C$.

Найдем вероятность появления контейнера $\forall x \in \boldsymbol{X}$:

$$\forall x_C \in \boldsymbol{X} \Rightarrow P_C(x_C) = E_C(e_C) \cdot Y_C(y_C); \tag{6}$$

$$\forall x_S \in \boldsymbol{X} \Rightarrow P_S(x_S) = E_C(e_S) \cdot Y_C(y_S). \tag{7}$$

Из условий леммы и формул (5)–(7) для каждой пары $x_S = x_C$ следует, что

$$x_S = x_C \Rightarrow P_S\left(x_S\right) = P_C\left(x_C\right). \tag{8}$$

Из утверждения (8) легко видеть, что

$$x_S = x_C \Rightarrow \log_2 \frac{P_C\left(x_C\right)}{P_S\left(x_S\right)} = \log_2 1 = 0. \tag{9}$$

Подставляя результат формулы (9) в определение относительной энтропии (см. формулу (1)), получим, что $D\left(P_C \middle\| P_S\right) = 0$. Следовательно, стегосистема является *совершенной системой (от пассивного противника)*. ►

Замечание. Условие леммы о декодировании помехоустойчивого кода с почти достоверной вероятностью и условие, что $Y_C\left(y\right) = Y_S\left(y\right)$, вполне естественны. В дальнейшем будем считать, что эти условия всегда выполняются и что любая *идеальная стегосистема* является *совершенной*. (Заметим, что обратное утверждение неверно — не любая совершенная стегосистема будет идеальной.) Таким образом, «идеальность» является достаточным условием для «совершенности» стеганографической системы.

Пример математической модели возникновения ошибок. Рассмотрим математическую модель (ММ) возникновения ошибок, представленную на рис. 3. Эта модель разработана нами при проектировании аппаратно-программного решения, реализующего стеганографию в кодах исправления ошибок на оптических дисках. Собранных нами статистических данных пока недостаточно, чтобы говорить об адекватности рассматриваемой ММ для оптических дисков. Однако в рамках настоящей статьи воспользуемся этой моделью в качестве примера реализации совершенной стеганографической системы.

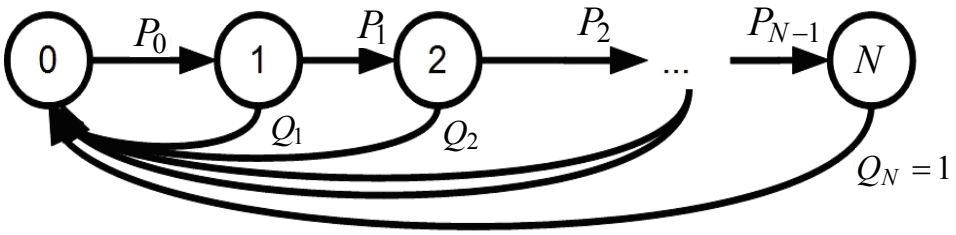

Рис. 3. Математическая модель возникновения ошибок на оптическом диске

При чтении уже записанного диска система находится в состоянии 0. Ошибка произойдет с вероятностью P_0. Если ошибка произо-

шла, то с вероятностью P_1 она продолжится. В общем случае если произошла ошибка P_i $(i \le N-2)$, то с вероятностью P_{i+1} она продолжится. Вероятности Q_i (вероятность того, что ошибки не произойдет) вычисляются как $Q_i = 1 - P_i$. Данная ММ представляет собой *марковский процесс* [4].

Обозначим через A_W вектор ошибок, произошедших на одном носителе данных емкостью *W* бит. Емкость можно измерять не только в битах, но и в байтах, буквах и т. п.; далее будем измерять в битах. Согласно ММ, приведенной на рис. 3, в случае возникновения ошибки длиной *m* необходимо, чтобы после ошибки был хотя бы один бит отсутствия ошибки. Таким образом, зная вектор A_W, можно вычислить количество произошедших ошибок каждой длины. Например, если $A_{16} = 0011001100111100$, то произошли две ошибки длиной 2 и одна ошибка длиной 4.

Введем обозначения:

m_i — математическое ожидание ошибок длиной *i* на одном носителе определенного производителя из определенной партии;

M — математическое ожидание общего количества ошибок любой длины;

$P(\xi = i \mid P_0)$ — вероятность того, что длина ошибки будет равна *i*, при условии, что произошла ошибка.

Тогда для ММ, изображенной на рис. 3, верны следующие формулы:

$$\forall i = \overline{1, N-1} \Rightarrow P(\xi = i \mid P_0) = (1 - P_i) \prod_{j=1}^{i-1} P_j , \tag{10}$$

$$P(\xi = N \mid P_0) = \prod_{j=1}^{N-1} P_j . \tag{11}$$

С помощью помехоустойчивого кода можно найти A_W для носителя данных. Для A_W можно подсчитать количество ошибок каждой длины. Пусть $A_W(j)$ есть A_W для *j*-го носителя.

Возьмем *k* носителей и $\forall j \in \{1,...,k\}$ найдем $A_W(j)$ и посчитаем количество ошибок каждой длины.

Обозначим через $m(j,i)$ количество ошибок длиной *i* на *j*-м носителе, а через $M(j)$ количество всех ошибок на *j*-м носителе. Можно легко увидеть, что:

$$M(j)=\sum_{i=1}^{N} m(j,i), \tag{12}$$

$$m_i=\lim_{k\to\infty}\frac{\sum_{j=1}^{k} m(j,i)}{k}, \tag{13}$$

$$M=\lim_{k\to\infty}\frac{\sum_{j=1}^{k} M(j)}{k}=\sum_{i=1}^{N} m_i, \tag{14}$$

$$P\left(\xi=i\middle|P_0\right)=\lim_{k\to\infty}\frac{1}{k}\cdot\sum_{j=1}^{k}\frac{m(j,i)}{M(j)}. \tag{15}$$

Вероятность возникновения ошибки найдем из следующих соображений: возьмем количество всех ошибок $M(j)$ и разделим на количество случаев, когда система находилась в состоянии 0 (см. рис. 3). Количество случаев, когда система находилась в состоянии 0, равно $W+M(j)-\sum_{i=1}^{N} i\cdot m(j,i)$. Если у нас *k* носителей, то необходимо взять среднее арифметическое. Таким образом, верна формула

$$P_0=\lim_{k\to\infty}\frac{1}{k}\sum_{j=1}^{k}\frac{M(j)}{W+M(j)-\sum_{i=1}^{N} i\cdot m(j,i)}. \tag{16}$$

На практике возьмем *k* достаточно большим. Выбор *k* должен зависеть от *W*. Если *W* очень большое (например, емкость оптического диска HVD [5]), то *k* можно принять за единицу.

По формуле (16) можно найти P_0. Затем с помощью формул (10), (11) и (15) можно найти сначала P_1, затем P_2 и т. д. Это означает, что для указанной математической модели можно статистическим способом найти вероятности P_i. Очевидно, что сложность данного алгоритма нахождения этих вероятностей $O(N)$.

В дальнейшем вероятности P_i будем представлять в виде *конечного ряда* и обозначать $\sum_{i=1}^{N} P_i$. Символ $\sum$ в данном случае обозначает ряд (т. е. упорядоченную последовательность чисел), а не сумму.

Совершенная стегосистема на основе ММ возникновения ошибок. Рассмотрим два различных носителя данных, ошибки которых можно задать с помощью ММ, приведенной на рис. 3. Пусть ряд $\sum_{i=1}^{N} P_i$ — это вероятности для первого типа носителей, а ряд $\sum_{i=1}^{N} R_i$ — вероятности для второго типа носителей. Предположим, что

$$\forall i = \overline{1, N-1} \Rightarrow P_i > R_i. \tag{17}$$

Физический смысл (17) таков: носители второго типа (и устройства записи для них) более качественные с точки зрения помех, чем носители первого типа. Например, в модели трех каналов первый ряд может характеризовать канал C_1, а второй ряд — канал C_2.

Возьмем *стегосообщение*, которое мы хотим вкрапить во второй, более качественный тип носителя. В модели трех каналов это будет канал C_2. Разобьем данное сообщение на последовательности длиной не более N так, чтобы конкатенация всех последовательностей образовывала искомое сообщение. Вкрапим каждую последовательность как искусственную ошибку в носитель. Подадим данный стегоконтейнер на вход в канал C_2. В итоге получим новое распределение с вероятностями $\hat{R}_i$.

Предположим теперь, что

$$\forall i = \overline{1, N-1} \Rightarrow P_i = \hat{R}_i. \tag{18}$$

Если справедлива формула (18), то распределения ошибок канала C_1, передающего пустые контейнеры, и канала C_2, передающего стегоконтейнеры, совпадут. Значит, мы получили *идеальную стегосистему* (по определению). Естественно предполагая, что код исправления ошибок с почти достоверной вероятностью декодирует искомую *информационную матрицу* и что $Y_{C_1}(y) = Y_{C_2}(y)$, по доказанной ранее лемме имеем *совершенную стегосистему*.

Формальный алгоритм вкрапления. Определение максимального объема вкрапляемого стегосообщения. Для вкрапления сообщения m в носитель данных необходимо знать ряды $\sum_{i=1}^{N} P_i$ и $\sum_{i=1}^{N} R_i$.

Определим ряд $\sum_{i=1}^{N} D_i = \sum_{i=1}^{N} (P_i - R_i)$. Пусть M — это разница количества ошибок любой длины между каналами C_1 и C_2 при передаче W бит данных контейнера.

Определим $D(\xi = i \mid D_0)$ аналогично формулам (10) и (11):

$$\forall i = \overline{1, N-1} \Rightarrow D(\xi = i \mid D_0) = (1 - D_i) \prod_{j=1}^{i-1} D_j , \tag{19}$$

$$D(\xi = N \mid D_0) = \prod_{j=1}^{N-1} D_j . \tag{20}$$

На практике определим натуральное μ_i следующим образом: $\mu_i = D(\xi = i \mid D_0) \cdot M$. Строго говоря, величина μ_i может не принадлежать натуральному множеству. При достаточно большом W данную величину можно округлить до натурального значения, при этом систему все равно будем считать совершенной. В теории для соблюдения математической строгости μ_i можно определить следующим способом:

$$\mu_i = \lfloor D(\xi = i \mid D_0) \cdot M \rfloor + \Delta(\mu_i), \tag{21}$$

где

$$\Delta(\mu_i) = \begin{cases} 1 & \text{с вероятностью } D(\xi = i \mid D_0) \cdot M - \lfloor D(\xi = i \mid D_0) \cdot M \rfloor, \\ 0 & \text{в противном случае.} \end{cases}$$

Разбиваем вкрапляемое сообщение на M различных последовательностей по правилу: из всех последовательностей ровно μ_i последовательностей длиной i. Вкрапляем каждую последовательность, начиная с определенной позиции контейнера, которая вычисляется генератором псевдослучайных чисел. Этот генератор однозначно определяет позицию по стегоключу и по порядковому номеру вкрапляемого сообщения (первое сообщение имеет номер 1, второе — 2, третье — 3 и т. д). Легко видеть, что по формуле (21) в среднем будем записывать $D(\xi = i \mid D_0) \cdot M$ последовательностей длиной i.

Максимальная длина вкрапляемого сообщения L вычисляется по формуле

$$L = \sum_{i=1}^{N} i\mu_i . \tag{22}$$

Если длина вкрапляемого сообщения меньше L, то следует дописать сообщение случайным набором битов.

Будем обозначать количество всех последовательностей буквой l. Данную величину можно вычислить по формуле

$$l = \sum_{i=1}^{N} \mu_i \,. \tag{23}$$

Если длина сообщения больше L, будем считать, что данный алгоритм неприменим.

Замечание. Существует вероятность, что подлинная ошибка попадет на искусственную (содержащую передаваемое скрытое сообщения). Для того чтобы можно было извлечь стегосообщение, необходимо перед вкраплением подать данное стегосообщение в кодер исправления ошибок. Предположим, что скорость кода для исправления ошибок, который используется для кодирования стегосообщения, равна $r < 1$. Тогда максимальный объем передаваемых скрытых данных не должен превышать Lr бит на один носитель емкостью
W бит.

Вероятность идеальности для рассматриваемой стеганографической системы. Теперь необходимо определить, является ли рассматриваемая стегосистема идеальной. Если *подлинная ошибка* не попала на место *искусственной ошибки*, то получим идеальную, а значит, и совершенную стегосистему. Обозначим вероятность этого события $P_{\text{ид}}$ и будем называть ее *вероятностью идеальности*.

Возьмем множество всех *искусственных ошибок* и зафиксируем их.

Пусть P_1 — вероятность того, что подлинная ошибка не началась в битах искусственной ошибки, а P_2 — вероятность того, что ошибка, начатая не на искусственной ошибке, «дошла» до бита искусственной ошибки. Например, искусственная ошибка располагается с 5-го по 10-й бит, а подлинная — началась с 1-го бита и закончилась на 7-м, т. е. подлинная ошибка началась не на искусственной, но «успела задеть» два бита искусственной ошибки.

Таким образом, справедлива формула

$$P_{\text{ид}} = P_1 \cdot P_2 \,. \tag{24}$$

Зная, что всего записано μ_i искусственных ошибок длиной i, можно вычислить первый множитель:

$$P_1 = \prod_{i=1}^{N} (1 - P_0)^{\mu_i \cdot i} = (1 - P_0)^L \,. \tag{25}$$

Найдем границу снизу для первого множителя. Если перед каждой искусственной ошибкой любой длины в течение N бит не будет происходить событие P_0 (см. рис. 3), то поскольку длина ошибки не превышает N, никакая подлинная ошибка не попадет на биты искусственной ошибки. Следовательно, существует граница снизу:

$$P_2 > \prod_{i=1}^{N} \left(1-P_0\right)^{\mu_i N} = \left(1-P_0\right)^{lN} . \qquad (26)$$

Из формул (24)–(26) получаем нижнюю границу для $P_{\text{ид}}$:

$$P_{\text{ид}} > \left(1-P_0\right)^{L+lN} = P_{\text{н.гр.ид}} . \qquad (27)$$

Величину $P_{\text{н.гр.ид}}$ будем называть *нижней границей вероятности идеальности*.

При $P_0 = 10^{-7}$, $L = 102\,400$, $l = 10N$, $N = 200$ нижняя граница вероятности идеальности равна $0{,}9511152$. При $P_0 = 10^{-6}$ и тех же значениях L, l и N справедливо неравенство $P_{\text{ид}} > 0{,}605\,803$.

При $P_0 = 10^{-6}$, $L = 102\,400$, $l = 10N$, $N = 100$ вероятность идеальности больше $0{,}8177$.

Если $P_0 = 10^{-5}$, $L = 102\,400$, $l = 10N$, $N = 100$, то $P_{\text{ид}} > 0{,}13$. Как видим, даже при $P_0 = 10^{-5}$ граница снизу для вероятности идеальности хоть и мала, но далека от невозможного события. Иначе говоря, более чем в каждом десятом случае получим идеальную стеганографическую систему.

Более того, если на практике подлинная ошибка «недостаточно часто» будет попадать на искусственную, то данная система, возможно, будет ε-*секретной стегосистемой* [1], где ε близка к нулю. На момент публикации статьи этот вопрос остается открытым.

Статистический подсчет вероятности. Можно подсчитать $P_{\text{ид}}$ на практике. Для этого следует записать U стегоконтейнеров, подать их на вход канала C_2, а на выходе из него посмотреть, попала ли подлинная ошибка на искусственную. Пусть в u случаях она не попала на искусственную, тогда в пределе

$$P_{\text{ид}} = \lim_{U\to\infty} \frac{u}{U} . \qquad (28)$$

При достаточно большом U можно принять $P_{\text{ид}} \approx \frac{u}{U}$.

Выводы:

1. Любая *идеальная стегосистема* для канала C_2 по отношению к каналу C_1 при корректности декодирования кодом исправления ошибок и при $Y_{C_1}\left(y\right) = Y_{C_2}\left(y\right)$ всегда является *совершенной стегосистемой*.

2. Стойкость совершенной стегосистемы при ее идеальности обусловлена корректностью математической модели возникновения ошибок. Если ММ адекватна, то система *совершенная.* Если существует более точная модель возникновения ошибок и противник знает о ней, а также может вычислить параметры ММ, то в рамках ММ противника система *не является совершенной*. Возможность вычисления параметров ММ — это необходимое условие, так как сложность вычисления параметров может быть настолько высокой, что ММ с практической точки зрения будет неприменимой, несмотря на большую точность. В примере, приведенном в данной статье, параметрами ММ являются величины P_i, и они вычисляются за $O(N)$. Таким образом, говоря о совершенной стегосистеме, мы должны считать адекватной саму математическую модель возникновения ошибок в канале передачи данных.

3. Для любой математической модели возникновения ошибок, рассчитав ее параметры, можно попытаться построить стеганографический алгоритм, который позволял бы реализовать *идеальную* (а значит, и *совершенную*) стегосистему. Для этого необходимо, чтобы параметры ММ возникновения ошибок в канале C_2 при передаче стегоконтейнеров совпадали с параметрами ММ возникновения ошибок в канале C_1 при передаче по нему пустых контейнеров. Затем необходимо найти вероятность идеальности. Если вероятность идеальности *близка к единице*, то на практике будем считать стегосистему *совершенной*.

4. Если нижняя граница вероятности идеальности достаточно высока, то можно считать, что на практике вероятность попадания подлинной ошибки на искусственную не изменяет существенно распределение ошибок, т. е. при построении стегосистемы можно принять эту вероятность за ноль и считать описанный алгоритм вкрапления *идеальной стегосистемой*. Это справедливо, так как на практике вероятность появления ошибки мала.

ЛИТЕРАТУРА

[1] Cachin C. An Information-Theoretic Model for Steganography. 2nd ed. *MIT Laboratory for Computer Science* — 2002, October, 2010, p. 31.

[2] Слипенчук П.В. Стеганография в кодах, исправляющих ошибки. *Вестник МГТУ им. Н.Э. Баумана. Сер. Приборостроение,* спец. вып. № 5, 2012, с. 249–260.

[3] Shannon C.E. A Mathematical Theory of Communication. *Bell System Technical Journal*, 1948, p. 623.

[4] Волков И.К., Зуев С.М., Цветкова Г.М. *Случайные процессы*, 3-е изд. Москва, МГТУ им. Н.Э. Баумана, 2006, с. 163–190.

[5] *International Standard ECMA-377. Information Interchange on Holographic Versatile Disc (HVD) Recordable Cartridges – Capacity: 200 Gbytes per Cartridge*. May, 2007.

Статья поступила в редакцию 28.06.2013

Ссылку на эту статью просим оформлять следующим образом:

Слипенчук П.В. Простое построение совершенных стегосистем на основе различных ошибок в помехоустойчивых кодах в модели трех каналов. *Инженерный журнал: наука и инновации,* 2013, вып. 11. URL: http://engjournal.ru/catalog/it/security/998.html

Слипенчук Павел Владимирович родился в 1990 г., окончил физико-математический лицей № 239 в Санкт-Петербурге. Студент кафедры «Информационная безопасность» МГТУ им. Н.Э. Баумана. e-mail: PVSlipenchoock@yandex.ru