P. V. Slipenchuk

Principles and Perspectives of Steganography in Error Correction Codes

Keywords: steganography in error correction codes, steganography goals, model of three channels

The main preference of a steganography in ECC with other types of steganography are listed. Opportunity and relevance of strict mathematical researches for this type of steganography are explained. The goals of a steganography in ECC are listed.

П. В. Слипенчук

ПЕРСПЕКТИВЫ И ПРАКТИЧЕСКОЕ ПРИМЕНЕНИЕ СТЕГАНОГРАФИИ В ПОМЕХОУСТОЙЧИВЫХ КОДАХ

Введение

Стеганография — это искусство и наука передавать сообщения различными способами так, чтобы не было обнаружено наличие самого сообщения, это область знаний о сокрытии информации; это процесс вкрапления представленной в какой-либо форме информации внутрь другой информации.

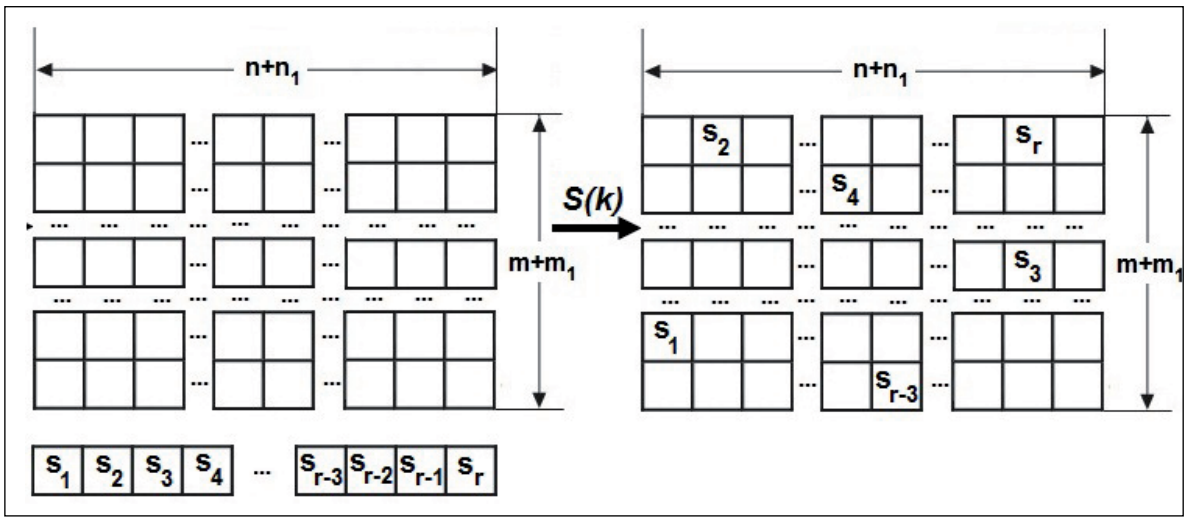

Рис. 1. Принцип стеганографии в помехоустойчивых кодах

Принцип **стеганографии в помехоустойчивых кодах** можно найти в работах [8, 9]. В общем случае есть блок, состоящий из букв какого-либо алфавита (например, из {0,1}). Определим размеры блока n на m. С помощью помехоустойчивого кода A мы преобразуем блок размера n на m в блок, размера $n+n_1$ на $m+m_1$ (см. рис. 1). В частном случае $m= m+m_1= 1$.

Блок, подаваемый на вход алгоритма, будем называть **информационным вектором** (или **информационной матрицей**), а выход алгоритма A будем называть **кодовым вектором** (или **кодовой матрицей**) [8].

Будем считать, что мы исследовали канал и нам известно распределение ошибок в канале. Имея данное конкретное распределение, можно построить алгоритм **S**, зависящий в общем случае от стегоключа **k**, который принимает на вход *кодовую матрицу (пустой контейнер)* и *стего-сообщение*, а на выходе дает стегоконтейнер в виде исходной кодовой матрицы с «вкрапленными» на определенных позициях битами сообщения (см. рис. 1).

Эти вкрапленные биты будем называть **искусственными ошибками** (artificial errors).

Затем полученный стегоконтейнер передаем по каналу с шумом. Ошибки, которые происходят в канале передачи данных, будем называть **подлинными ошибками** (genuine errors). Легко видеть, что по причине наличия не только *подлинных*, но и *искусственных* ошибок общее количество ошибок в канале изменится.

1. Практическое применение стеганографии в кодах, исправляющих ошибки

С практической точки зрения, любая стеганография (необязательно в помехоустойчивых кодах) в общем случае может быть использована только в трех целях:

1) Цифровые отпечатки (ЦО) (Digital Fingerprint). Данный вид стеганографии предназначен для того, чтобы можно было обнаружить, скопирован контейнер или нет. Например, в каждую продаваемую копию электронной книги можно вкраплять скрытое сообщение. Если пользователь, купивший экземпляр электронной книги, опубликует его на каком-нибудь интернет-ресурсе, то с помощью ЦО можно будет идентифицировать злоумышленника. Таким образом ЦО могут быть применимы для защиты исключительного права. Если же с помощью какого-либо алгоритма третья сторона сможет извлечь ЦО из контейнера, то идентифицировать злоумышленника невозможно, но до тех пор, пока третья сторона не научится подделывать ЦО, она не сможет безнаказанно продавать электронную копию книги. Таким образом, при извлечении ЦО третья сторона может преследовать две цели:

а) извлечение ЦО из контейнера;

б) подмена одного ЦО другим ЦО.

Первую задачу будем называть **задачей устранения** ЦО, а вторую — **задачей подмены** ЦО.

2) Стеганографический водяной знак (СВЗ) (Stego Watermarking). Задача водяного знака — подтвердить авторскую подлинность защищаемого контейнера. Например, при съемках на видеокамеру можно в каждый кадр вкраплять информацию о времени записи, модели видеокамеры или имени оператора видеокамеры. В случае если отснятый материал попадет в руки конкурирующей компании, вы можете попытаться использовать СВЗ для подтверждения авторства записи на суде. Таким образом, СВЗ может быть применим для защиты права авторства.

3) Скрытая передача данных (СПД). Этот вид стеганографии в основном используется в военных и криминальных целях для незаметной передачи данных из одной точки в другую. Например, дипломат одного государства может попытаться выкрасть чертежи модели баллистических ракет и спрятать их во внешне невинный аудиофайл, или менеджер одной крупной компании может украсть бухгалтерские отчеты и вкрапить их в фотографию.

В современных статьях, посвященных стеганографии, часто используется термин **ЦВЗ (Цифровые водяные знаки)**. Под этим термином подразумевают то СВЗ, то ЦО. Тем не менее задачи и проблемы при реализации ЦО и СВЗ принципиально различные! Действительно, СВЗ на всех копиях электронного документа одинаков, а ЦО на всех копиях документов различен.

Стеганография в помехоустойчивых кодах применима для реализации ЦО и скрытой передачи данных, но не применима для СВЗ: так как задача СВЗ — подтвердить авторство электронного документа, то исправление ошибок контейнера уничтожит СВЗ. Таким образом, стеганография в кодах, исправляющих ошибки, принципиально неприменима в качестве защиты *права авторства* (СВЗ), но применима для защиты *исключительного прав*а (ЦО). *Задача устранения* ЦО для помехоустойчивых кодов решается тривиально — следует исправить все ошибки. Но *задача подмены* может быть нетривиальной.

2. Модель Кашена и совершенно секретная стегосистема

В работе [1] Кристиан Кашен предлагает стеганографическую модель:

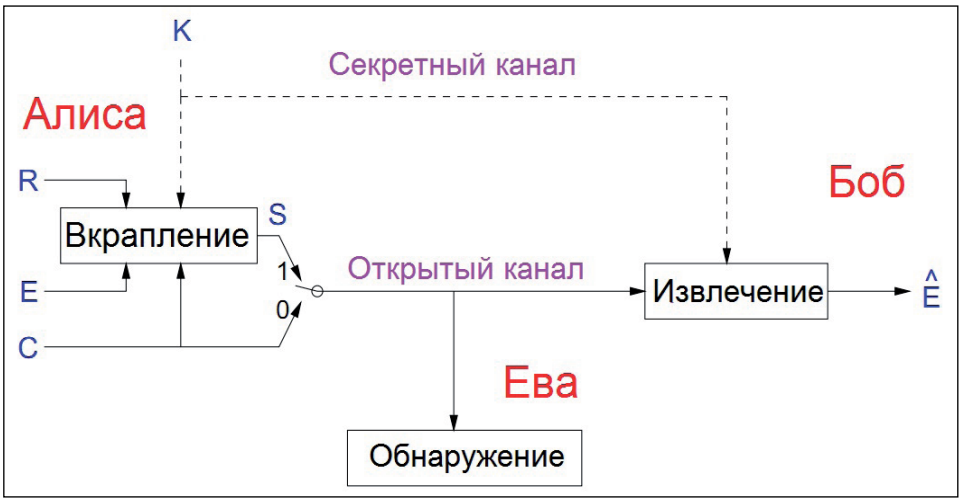

Рис. 2. Общая стеганографическая модель Кристиана Кашена

На рис. 2 показаны операции стегосистемы. Переключатель **S** определяет состояние Алисы:

- Первое состояние называется пассивным (переключатель в позиции 0). Алиса отправляет только пустые контейнеры **C** Бобу по открытому каналу передачи данных.
- Второе состояние называется активным (переключатель в позиции 1). В данном случае Алиса отправляет *стегосообщение* **E**, которые она вкрапила в пустой контейнер **C**, используя алгоритм **A**. Алгоритм **A** берет на вход **C**, ключ **K**, закрытый случайный источник **R**, сообщение **E** и выдает на выходе *стегоконтейнер* **S**. Стегоконтейнер отправляется в открытый канал Бобу. Противник Ева и получатель Боб принимают **S**. Используя *алгоритм извлечения* **B**, Боб извлекает сообщение **E** из **S** с помощью **K** в надежде, что он получил скрытое сообщение от Алисы.

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

Определим величину $D(P_{X1} \| P_{X1})$ как:

$$D(P_{X1} \| P_{X2}) = \sum_{x \in X} P_{X1}(x) \log \frac{P_{X1}(x)}{P_{X2}(x)} . \quad (1)$$

Определим множество $\{\hat{\mathbb{R}} = \mathbb{R} \cup \{\infty\}\}$. Величину $D(P_{X2} \| P_{X2})$, заданную формулой (1), при условии, что X1 — это множество X, а X2 — это множество $\hat{\mathbb{R}}$, будем называть **относительной энтропией (relative entropy)**[1], если определить $0 \log \frac{0}{0} = 0$; $T \log \frac{T}{0} = \infty$; и $D(P_{X1} \| P_{X2}) = \infty$, если хотя бы одно слагаемое равно ∞ .

Система называется **совершенно секретной (от пассивного противника),** если *относительная энтропия* между P_C и P_S равна нулю. Иначе говоря:

$$D(P_C \| P_S) = 0. \quad (2)$$

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

$$D(P_C \| P_S) \leq \epsilon . \quad (3)$$

[1] Обратим внимание, что относительная энтропия не является симметричной величиной! То есть $D(P_C \| P_S) \neq D(P_S \| P_C)$.

3. Модель трех каналов. Идеальная стегосистема

Модель трех каналов и идеальная стегосистема впервые определены в работе [8].

Рассмотрим три канала: C_1, C_2 и C_3. Будем считать, что по каким-либо характеристикам канал C_2 лучше канала C_1. Допустим, что стегоконтейнер будет проходить через канал с шумом C_2, а пустой контейнер — через канал с шумом C_1. Пусть **S** — стегосистема, вкладывающая сообщение определенной длины в пустой контейнер.

Рассмотрим схему передачи сообщения [8]:

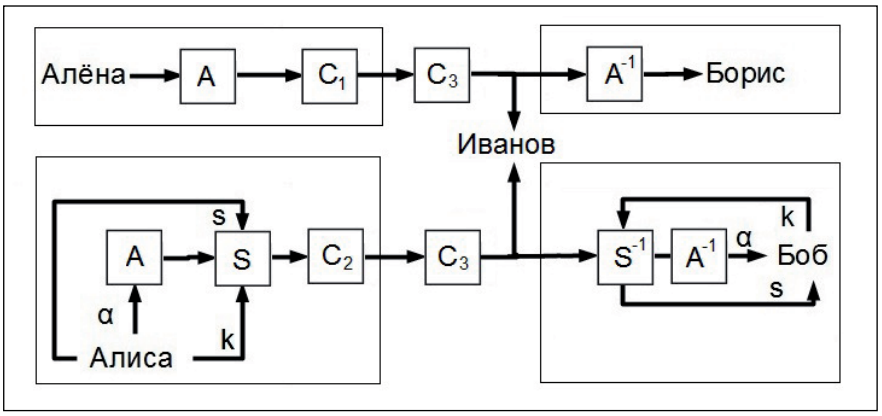

Рис. 3. Модель трех каналов

Алиса передает Бобу серию стегоконтейнеров. В общем случае Алиса берет информационную матрицу α, подает на вход кодеру **A** (см. рис. 3). Затем с помощью стегоалгоритма **S** и ключа **k** создает стегоконтейнер, положив в него некое сообщение **s**. Этот стегоконтейнер проходит через два канала с шумом: C_2 и C_3 (в частном случае канал C_3 — канал без шума). Боб принимает стегоконтейнер, с помощью ключа **k** извлекает стегосообщение **s** [8].

Алёна передает Борису пустые контейнеры. Они и не думают с помощью стеганографических алгоритмов передавать скрытые сообщения.

Иванов — третья сторона. Его задачи: 1) обнаружить, передается ли в контейнере скрытое сообщение; 2) если передается, то извлечь скрытое сообщение [8].

Рассмотрим *искусственные ошибки*, вкрапляемые в *кодовый вектор* перед каналом C_2. Саму ошибку будем обозначать буквой v. После прохождения стегоконтейнера через канал C_2 на выходе будем иметь стегоконтейнер, содержащий и *искусственные*, и *подлинные* ошибки. Ошибка из *искусственных* и *подлинных* ошибок представляет собой случайную величину e_2+v. Определим $x+y$ как побитную сумму по модулю два. Множество всех ошибок длины $(n+n_1)\cdot(m+m_1)$ будем обозначать как X. Для e_2+v справедлива формула возникновения ошибки [9]:

$$P(e_2+v=x)=\sum_{\forall y\in X} P(e_2=x+y)\cdot P(v=y) \quad (4)$$

Обозначим R_1 распределение случайной величины e_1, а 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 [8, 9].

4. Перспективы стеганографии в кодах, исправляющих ошибки, как строгой математической дисциплины

В работе [9] дано объяснение того, что модель трех каналов можно рассматривать как частный случай модели Кашена. Также в работе [9] есть доказательство того, что *идеальная* стегосистема

является *совершенно секретной (от пассивного противника)*. Таким образом, в случае стеганографии в помехоустойчивых кодах для доказательства *совершенной секретности* достаточно доказать *идеальность* стегосистемы. В работе [9] дан пример в виде математической модели возникновения ошибок на оптических дисках CD, способ нахождения параметров данной модели и перечислены условия, при которых стегосистему можно считать совершенно секретной.

Как уже было сказано в [9], основная проблема идеальных стегосистем заключена в том, каким именно способом рассчитываются R_1 и R_2. Эти величины рассчитываются с помощью некоторой математической модели ошибок. Но на основании чего мы принимаем, что разработанная математическая модель адекватна? Следовательно, проблема адекватности математической модели является основной задачей при обосновании совершенной секретности конкретной идеальной стегосистемы [9].

Проблема построения математической модели является частично проблемой математики, частично проблемой других наук. Но после построения математической модели ошибок мы имеем дело только с математикой. Поэтому стеганографию в помехоустойчивых кодах можно считать той областью стеганографии, в которой возможны строгие математические рассуждения.

Это, казалось бы, несущественное свойство имеет большое значение и представляет интерес для научного мира, так как сейчас среди специалистов в области криптографии существует законный скепсис по отношению к стеганографии. Это вызвано тем, что многие методы обнаружения скрытого сообщения в различных типах стеганографии основаны на экспертном анализе, то есть на личном опыте и интуиции. Автор не хочет обидеть экспертный метод анализа. Просто его использование не позволяет назвать ту или иную дисциплину, область знаний и исследований математической наукой.

5. Преимущества стеганографии в кодах, исправляющих ошибки

Стеганография в помехоустойчивых кодах имеет ряд преимуществ по сравнению с другими видами стеганографии.

1) Независимость от данных контейнера. Стеганография в кодах для исправления ошибок не зависит от содержания данных в канале[2], что нельзя сказать про другие методы стеганографии. К примеру, при методе LSB (Least Significant Bit) сообщение вкрапляется в наименее значащие биты изображения. Подобным образом нельзя поступить с текстом, так как изменение любой буквы в файле будет заведомо рассмотрено как опечатка или ошибка. Стеганография в кодах для исправления ошибок универсальна и не зависит от того, какие данные представляет собой *информационный вектор*, поданный на вход помехоустойчивого кода.

2) **Возможно построение строгой математической модели.** Как уже было сказано ранее, для исправляющих кодов в оптических дисках можно построить строгую математическую модель возникновения ошибок, а значит, данный вид стеганографии можно считать строгой математической наукой.

3) **Распространенность контейнеров.** Помехоустойчивые коды применяются практически во всех способах передачи информации. Это Wi-Fi, спутниковая связь, твердотельный накопитель (Solid-State Drive, SSD), накопитель на жестких магнитных дисках (НЖМД, hard disk drive, HDD), флеш-накопители и т. д.

4) **Техническая сложность стегоанализа.** Как правило, процесс записи и извлечения информации на уровне помехоустойчивых кодов требует специального оборудования и/или очень много времени. Связано это с тем, что данные преобразования выполняются, как правило, на микроконтроллере периферийного оборудования или на заказной микросхеме. Предоставление пользователю оборудования возможности программного чтения и записи помехоустойчивых кодов не является основным функционалом, поэтому такой возможности или просто нет, или она является факультативной.

[2] Или на носителе данных. В дальнейшем будем говорить «в канале», подразумевая также и «носители», например оптические диски.

Например, для CD-дисков можно извлечь данные только RSPC-кодов [2] с помощью факультативной команды “READ CD” [3], которая для большинства современных оптических приводов редко поддерживается. Коды CIRC и RLL [2] извлечь с диска программными средствами невозможно. Более того, на большинстве современных оптических приводов не поддерживается эта команда в режиме чтения RSPC[3]. Для других популярных форматов оптических носителей, таких как DVD [4], BD [7] и HVD [5, 6], не предусмотрено даже факультативных команд извлечения *кодовых векторов*. Таким образом, пользователь может скопировать данные, но не может извлечь стегоконтейнер без специального оборудования.

5) **Цифровые отпечатки.** Стеганография в кодах для исправления ошибок на лицензионных носителях может быть хорошим способом для защиты *исключительного права*. Например, можно каждый лицензионный оптический диск помечать определенной информацией. Места вкрапления и сама информация может быть уникальна для каждого диска и определяться по ключу. Допустим, злоумышленник попробует скопировать диск с целью дальнейшей перепродажи. Тогда возможно следующее:

а) противник скопирует данные на высоком уровне и потеряет цифровой отпечаток;

б) противник скопирует данные на низком уровне, но тогда у него будут копии с одним и тем же ЦО. Если проверочная комиссия обнаружит хотя бы два диска с одинаковой подписью (или хотя бы один с отсутствием ЦО), то это может служить доказательством факта незаконного копирования.

6) **Специальное использование.** С развитием стеганографии как науки ее использование расширится: стеганография станет своего рода «информационным противостоянием» между полицией и организованной преступностью, между различными государствами; данный тип стеганографии может быть использован как канал скрытой передачи данных (СПД).

Выводы

1. Для стеганографии в помехоустойчивых кодах возможны математические исследования, чего, к сожалению, пока нельзя сказать о большинстве стеганографических приемов.

2. Стеганография в помехоустойчивых кодах имеет ряд преимуществ по сравнению с другими типами стеганографии. В статье дано их описание.

3. Контейнеры для стеганографии в помехоустойчивых кодах распространены. Изучение данного типа стеганографии актуально и может быть востребовано как в специальном (СПД), так и в гражданском (ЦО) применении.

4. Стеганография в помехоустойчивых кодах не может быть применена для реализации стеганографических водяных знаков (СВЗ).

СПИСОК ЛИТЕРАТУРЫ:

1. *Cachin C.* An Information-Theoretic Model for Steganography // MIT Laboratory for Computer Science — 2002.— 31 p.
2. International Standard ECMA-130. Data interchange on read-only 120 mm optical data disks (CD-ROM). June. 1996.
3. International Standard MMC-6 (Multi-Media Commands - 6) // International Committee for Information Technology Standards. Revision 2. 9. July. 2008.
4. International Standard ECMA-338. 80 mm (1,46 Gbytes per side) and 120 mm (4,70 Gbytes per side) DVD Re-recordable Disk (DVD-RW). December. 2002.
5. International Standard ECMA-377. Information Interchange on Holographic Versatile Disc (HVD) Recordable Cartridges — Capacity: 200 Gbytes per Cartridge.1st Edition. May. 2007. Секция 14. Глава 2. Параграф 2.3
6. International Standard ECMA-378. Information Interchange on Read-Only Memory Holographic Versatile Disc (HVD-ROM) — Capacity 100 Gbytes per disk. 1st Edition. May. 2007.
7. Standard Blu-ray Disk Format: 1.A Physical Format Specifications for BD-RE. 3rd Edition. October. 2010.

[3] Не путать с командой READ, которая факультативной не является, в отличие от команды READ CD.

8. *Слипенчук П. В.* Стеганография в кодах, исправляющих ошибки // Вестник МГТУ. Специальный выпуск № 5. 2013
9. *Слипенчук П. В.* Простое построение совершенных стегосистем на основе различных ошибок в помехоустойчивых кодах в модели трех каналов // Вестник МГТУ. Специальный выпуск (в печати).

REFERENCES:

1. *Cachin C.* An Information-Theoretic Model for Steganography // MIT Laboratory for Computer Science - 2002. Edition. October, 2010. P. 31.
2. International Standard ECMA-130. Data interchange on read-only 120 mm optical data disks (CD-ROM). June 1996.
3. International Standard MMC-6 (Multi-Media Commands - 6) // International Committee for Information Technology Standards. Revision 2, 9 July 2008.
4. International Standard ECMA-338. 80 mm (1,46 Gbytes per side) and 120 mm (4,70 Gbytes per side) DVD Re-recordable Disk (DVD-RW). December 2002.
5. International Standard ECMA-377. Information Interchange on Holographic Versatile Disc (HVD) Recordable Cartridges — Capacity: 200 Gbytes per Cartridge.1st Edition. May, 2007. 14.2.2.3
6. International Standard ECMA-378. Information Interchange on Read-Only Memory Holographic Versatile Disc (HVD-ROM) - Capacity 100 Gbytes per disk. 1st Edition. May, 2007.
7.Standard Blu-ray Disk Format: 1.A Physical Format Specifications for BD-RE. 3rd Edition. October, 2010.
8. *Slipenchuk P. V.* Steganografiay v kodakh, ispravlayushikh oshibri // M.: Vestnik MGTU, 2013. Spetsialniy Vipusk № 5.
9. *Slipenchuk P. V.* Prostoe postroenie sovershennikh stegosistem na osnove razlichnikh oshibok v pomekhoustoiychivikh kodakh v modeli trekh kanalov // M.: Vesnik MGTU, Cpetsialniy vipusk (v pechati).