УДК 621.391

П.В. Слипенчук

СТЕГАНОГРАФИЯ В КОДАХ, ИСПРАВЛЯЮЩИХ ОШИБКИ

Проанализирован ряд практических и теоретических проблем, возникающих при использовании стеганографии в кодах, исправляющих ошибки (Error Correction Code, ECC). Дан краткий анализ хорошо известных стеганографических приемов и методов. Изложены новые идеи применения стеганографических методов в кодах, исправляющих ошибки. Обоснована возможность и перспективность реализации стеганографических алгоритмов в кодах, исправляющих ошибки в цифровых многоцелевых дисках (Digital Versatile Disc, DVD), дисках Blu-Ray (BD) и голографических многоцелевых дисках (Holographic Versatile Disc, HVD).

E-mail: PVSlipenchoock@yandex.ru

***Ключевые слова*:** *стеганографические модели, стеганография в кодах, исправляющих ошибки, DVD, BD, HVD.*

Стеганографические системы. Стеганографические системы можно условно подразделить на ***физические*** (симпатические чернила, запись внутри вареного яйца, микроточки и т. д.) и ***информационные системы***. Физическая стеганография подразумевает использование свойств какого-либо объекта, на которые человек не обращает должного внимания. Эти свойства и являются скрытым каналом передачи информации. В информационной стеганографии сокрытие данных происходит непосредственно внутри слов какого-либо алфавита, например алфавита букв естественного языка или алфавита 0,1. Представление битового потока (совокупности слов) на реальном физическом носителе неважно.

В статье будут использованы следующие основные определения из области стеганографии [1].

Сообщение (или стегосообщение) — передаваемая (или хранимая) скрытая информация.

Контейнер — любая информация, используемая для сокрытия сообщения.

Пустой контейнер — контейнер, не содержащий сообщения.

Заполненный контейнер (стегоконтейнер) — контейнер, содержащий сообщение.

Ключ (стегоключ) — секретный параметр, необходимый для сокрытия сообщения в контейнере.

Стеганографический канал (стегоканал) — канал передачи стегоконтейнера.

Под ***стеганографической системой (стегосистемой)*** будет пониматься программная, аппаратная или программно-аппаратная система, пригодная для организации скрытого канала передачи информации. Аналогично криптографическому принципу Кирхгофа,

полагаем, что третья сторона точно знает алгоритм работы стеганографической системы. Неизвестным для третьей стороны остается только секретный ключ, с помощью которого можно узнать о факте существования сообщения и о его содержании. Таким образом, при проектировании стеганографической системы следует руководствоваться следующими принципами:

а) без знания ключа третьей стороне должно быть затруднительно даже установить факт существования сообщения;

б) при обнаружении противником наличия скрытого сообщения он не должен иметь возможности извлечь сообщение до тех пор, пока не будет владеть ключом.

Задача стегоанализа сводится к подзадачам:

— обнаружения факта наличия сообщения;

— извлечения сообщения из заполненного контейнера.

Стеганографические системы можно подразделить на системы с ***произвольным контейнером*** и системы с ***подобранным контейнером***. Подобранный контейнер, в отличие от произвольного, строится непосредственно по конкретному ключу и конкретному сообщению. В случае с подобранным контейнером стегосистема принимает на вход сообщение и ключ, затем с помощью определенного алгоритма формирует стегоконтейнер.

Простым примером системы с подобранным контейнером может служить акростих Н. Гумилева [2]:

Аддис-Абеба, город роз.
На берегу ручьев прозрачных,
Небесный див тебя принес,
Алмазной, средь ущелий мрачных.

Армидин сад… Там пилигрим
Хранит обет любви неясной
(Мы все склоняемся пред ним),
А розы душны, розы красны.

Там смотрит в душу чей-то взор,
Отравы полный и обманов,
В садах высоких сикомор,
Аллеях сумрачных платанов.

Если взять первые буквы каждой строки, то получится имя супруги Гумилева: АННА АХМАТОВА. В данном случае стеганографическая система извлечения сообщения при принятии на вход ключа должна указать позиции, с которых необходимо брать символы сообщения (в случае акростиха Гумилева — каждая первая буква строки, кроме третьей строки второго четверостишия, если скобку считать буквой).

На основе акростихов в стеганографии используют метод ***подражательных функций*** (Mimic Functions). Подробнее о подража-

тельных функциях можно прочитать в работе [3]. Однако такой метод стеганографии имеет слабую производительность и небольшую долю сообщений в контейнере [4].

На вход системы с произвольным контейнером подается стегоключ, сообщение и пустой контейнер. На выходе система выдает заполненный контейнер. Эти системы можно подразделить на системы ***с потерей информации (в контейнере)*** и системы ***без потери информации.***

В системах с потерей информации информация пустого контейнера обладает большой избыточностью. К подобным контейнерам относятся: аудиофайлы, файлы-изображения и их комбинации в виде видеофайлов. Для таких контейнеров вследствие избыточности успешно используются алгоритмы сжатия без потерь (например, tar, zip, rar), а также сжатия с потерями (например, JPEG, JPEG-2000, MPEG). Однако системы с потерей информации в ближайшем будущем могут перестать применяться. Действительно, если после сжатия контейнера из него по-прежнему можно извлечь сообщение, то почему бы не исключить из контейнера биты, используемые для восстановления сообщения?

Для более формального обоснования изложенного выше используем следующие определения: стегосистема A называется ***робастной к алгоритму сжатия*** B, если существует алгоритм извлечения, позволяющий всегда восстановить сообщение после применения к любому контейнеру системы A алгоритма сжатия B. В противном случае систему A назовем ***хрупкой к алгоритму сжатия*** B. Если алгоритм робастен к алгоритму сжатия B, то это свидетельствует о том, что алгоритм B «не до конца хорошо» сжимает контейнер. Если после сжатия полученный контейнер можно принять ***«неотличимым»*** (например, для глаза человека) и в нем есть биты скрытого сообщения, то почему бы не сжать контейнер «ещё сильнее», удалив из него те самые биты, которые могут использоваться для скрытого сообщения? Таким образом, принципы построения и использования робастных систем основаны на применении недостаточно эффективных алгоритмов сжатия. При развитии алгоритмов сжатия системы станут хрупкими (если, конечно, при проектировании нового алгоритма не ставится задача стеганографии).

Рассмотрим алгоритмы сжатия без потерь. Существует точка зрения, что при вложении сообщения в пустой контейнер, энтропия по Шенону контейнера увеличивается, и соответственно при сжатии контейнера в среднем получается архив большего объема, чем архив из пустого контейнера [5]. Это очень похоже на правду, но строгих математических доказательств приведенного утверждения неизвестно.

Системы без потери информации включают в себя ***файловую стеганографию*** и ***стеганографию в кодах, исправляющих ошибки*** (Error Correction Code, ECC). Файловая стеганография основана на записи сообщения в редко используемые биты файлов (заголовки

протоколов, области данных и т. п.), или в биты файлов, которые никогда не применяются (например, зарезервированные). Однако при элементарном анализе в любых файлах сразу устанавливается наличие скрываемых в этих местах данных [4].

Более интересно рассмотреть стеганографию в освобожденных участках энергонезависимых запоминающих устройств (например, накопитель на жестких магнитных дисках, НЖМД). Поскольку диск используется не полностью, можно в свободные части НЖМД записать сообщение. Разумеется, необходимо подумать про код, исправляющий ошибки. Так, возможна ситуация, когда в свободные для операционной системы участки памяти НЖМД записывается блок данных. Но эта же область НЖМД содержит часть стегосообщения. Тогда требуется продумать код, исправляющий ошибки, чтобы при записи в свободную память, в которой находится часть скрытого сообщения, других частей хватило для декодирования сообщения.

Особый интерес для файловой стеганографии представляет твердотельный накопитель (Solid-State Drive, SSD). В нем используется технология Flash-преобразования адресов (Flash Translation Layer, FTL). Вследствие быстрого износа при перезаписи блоков данных, каждый раз запись идет в разные физические адреса [6]. При доступе к памяти программы контроллера твердотельного накопителя можно реализовать эффективную стеганографическую систему.

Стеганография в кодах, исправляющих ошибки. В общем случае есть блок размерами $n \times m$, состоящий из букв какого-либо алфавита (например, из алфавита 0,1). С помощью кода $\boldsymbol{A}$, исправляющего ошибки, блок размерами $n \times m$ преобразуется в блок, размерами $(n + n_1) \times (m + m_1)$. В частном случае $m = m + m_1 = 1$ (рис. 1).

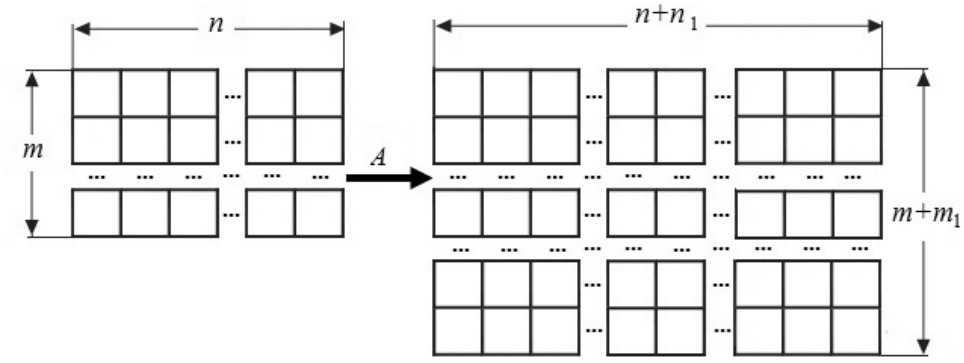

Рис. 1. Схема преобразования блока размерами $n \times m$ в блок размерами $(n + n_1) \times (m + m_1)$ с помощью кода, исправляющего ошибки

Блок, подаваемый на вход алгоритма, будем называть ***информационной матрицей***, а выход алгоритма A — ***кодовой матрицей***. Код, исправляющий ошибки, может быть создан для исправления одиночных или пакетных ошибок. Кодовая матрица поступает в ка-

нал с шумом, принимается второй стороной и декодируется в исходную информационную матрицу.

Допустим, известно распределение ошибок в канале. Анализируя конкретное распределение, можно построить алгоритм S, зависящий в общем случае от стегоключа k, который принимает на вход кодовую матрицу (пустой контейнер) и сообщение, а на выходе дает стегоконтейнер в виде исходной кодовой матрицы с «вкрапленными» на определенных позициях битами сообщения (рис. 2).

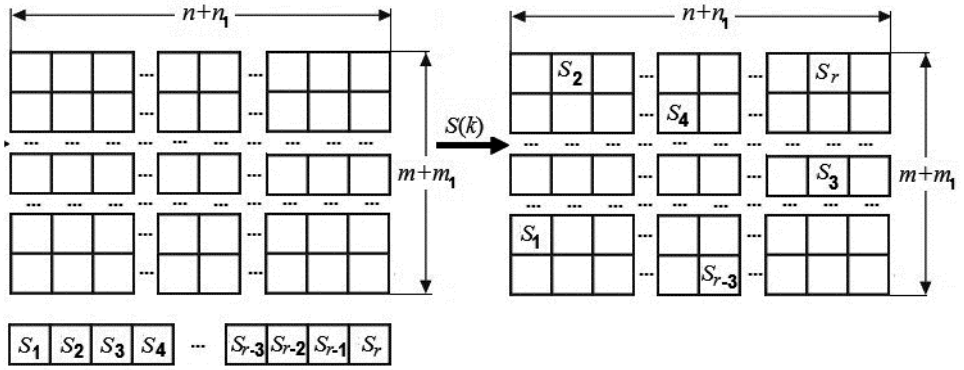

Рис. 2. Схема построения алгоритма *S*

Затем стегоконтейнер передается по каналу с шумом. Очевидно, что распределение возможных ошибок изменится по отношению к распределению ошибок при передаче пустого контейнера по каналу.

При такой системе требуется соблюдать два условия:

1) после «вкрапления» сообщения и прохождения контейнера через канал с шумом, необходимо, чтобы вероятность ошибки декодирования была по-прежнему крайне низкой;

2) распределение ошибок для контейнера должно быть максимально приближенным к распределению ошибок для пустого контейнера.

Первое условие необходимо для правильного восстановления исходной информационной матрицы, второе — для уменьшения «подозрительности» контейнера на наличие скрытого сообщения. Чем меньше байт затирает алгоритм $S(k)$, тем ближе распределения друг к другу и меньше объем скрываемой информации.

Следует учесть, что число бит, используемых алгоритмом $S(k)$, должно быть больше числа r (см. рис. 2). Это связано с тем, что ошибки в канале с шумом могут быть наложены на биты, содержащие скрытое сообщение (s_1, …, s_r), следовательно, скрытое сообщение также требуется пропустить через некий код, исправляющий ошибки.

Выше был рассмотрен один канал с шумом. По нему проходил и пустой контейнер и стегоконтейнер. При этом распределения ошибок при сообщении ненулевой длины должны быть разными.

Пусть даны два канала C_1 и C_2 с шумом, ошибок в канале C_2 в среднем меньше, чем в канале C_1. Допустим, что стегоконтейнер проходит через канал C_2 с шумом, а пустой контейнер через канал C_1 с шумом. Пусть S — стегосистема, вкладывающая определенной длины сообщение в пустой контейнер; R_1 и R_2 — распределения ошибок для пустого контейнера, проходящего через канал C_1, и для стегоконтейнера, проходящего через канал C_2 соответственно. 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.

Рассмотрим теперь следующую схему передачи сообщения (рис. 3). Алиса передает Бобу серию стегоконтейнеров. В общем случае Алиса берет информационную матрицу α и подает на вход кодеру A. Затем с помощью стегоалгоритма S и ключа k создает стегоконтейнер, введя в него некое сообщение s. Этот стегоконтейнер проходит через каналы с шумом C_2 и C_3 (в частном случае канал C_3 без шума). Боб получает стегоконтейнер и с помощью ключа k извлекает из него сообщение s.

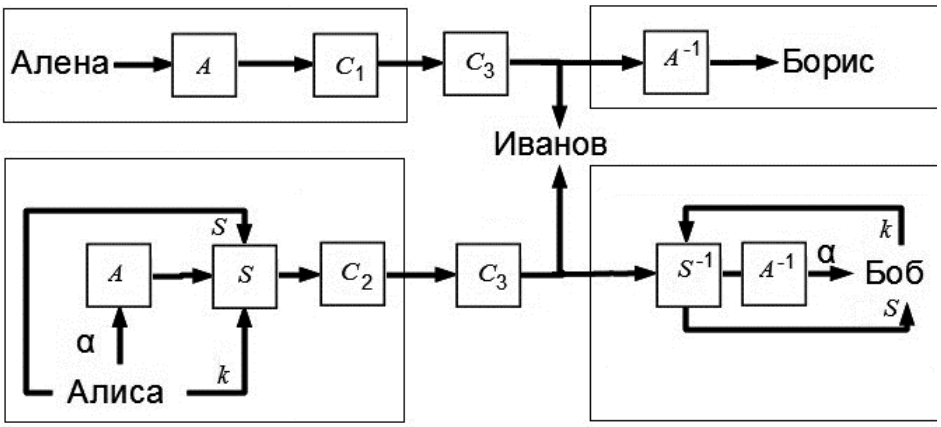

Рис. 3. Схема передачи сообщения

Алена передает Борису пустые контейнеры. При этом они не думают передавать скрытые сообщения с использованием стеганографических алгоритмов. Иванов — третья сторона. Его задачи: обнаружить передачу скрытого сообщения в контейнере; извлечь скрытое сообщение, если сообщение передается.

В качестве примера каналов можно привести процесс записи информации на цифровой многоцелевой диск (Digital Versatile Disc, DVD). Запись на него разным оборудованием с различной скоростью — это каналы C_1 и C_2. Передача диска или хранение на складе — канал C_3. В первом случае канал C_3 — канал без шума, во втором, при долгом и (или) неправильном сроке хранения — канал с шумом.

Предположим, что:

1) передается серия стегоконтейнеров (т. е. их число существенно больше, чем 1);

2) Иванов не знает не только ключа k, но и алгоритма S;

3) Иванов знает распределение ошибок в каналах C_1 и C_3;

4) Иванов может прослушивать сообщения после того, как они прошли канал C_3.

Поскольку Иванов может прослушивать сообщения канала C_3, то при большом числе передаваемых сообщений Иванов может вычислить распределение ошибок контейнеров от Алисы к Бобу и от Алены к Борису. Распределения ошибок контейнеров от Алены к Борису должны совпасть с теоретическими, так как они передают пустые контейнеры через каналы C_1 и C_3, распределение ошибок которых известно Иванову. Если у Алисы неидеальная система для канала C_2 по отношению к каналу C_1, то распределения ошибок контейнеров от Алисы к Бобу не совпадут. Чем «более существенно» отличаются распределения между каналами C_2 и C_1, тем больше вероятность корректного распознавания стегоконтейнера из множества контейнеров. Таким образом, Иванов может решить первую проблему стегоанализа. Однако для извлечения сообщения необходимо знать алгоритм S.

Отметим также, что каналы C_1 и C_2 могут быть более сложными, чем симметричный канал с шумом. Например, рассматриваемые каналы могут быть каналами с пакетами ошибок. В данном случае, Иванову мало знать распределение ошибок, он должен также учитывать распределение вероятностей длин пакетных ошибок.

Из изложенного выше следует, что при стеганографии в кодах, исправляющих ошибки, по схеме, приведенной на рис. 3, для обеспечения невозможности обнаружения скрытого сообщения необходимо, чтобы распределения ошибок после прохождения каналов C_1 и C_2 совпадали.

Стеганография в кодах, исправляющих ошибки оптических дисков. Данные, получаемые с ЭВМ и передаваемые на контроллер DVD-привода, называют ***основными данными*** (Main Data). Эти данные проходят обработку через несколько шагов, каждый из которых принимает данные с предыдущего шага, обрабатывает и передает данные на следующий шаг [7]:

1. Конфигурация таблицы данных (Data Frame) используется для разделения основных данных на так называемые таблицы данных, каждая из которых имеет свой уникальный идентификатор (Identification Data, ID), информацию о правах (Copyright Management Information, CPR MAI), поля обнаружения ошибок (Error Detection Code).

2. Скремблирование — побитовое сложение каждой таблицы данных с определенной гаммой и получение скремблированных таблиц (Scrambled Frame).

3. Конфигурация блоков кода, исправляющего ошибки (ECC Block) — каждые 16 скремблированных таблиц кодируются внешним

и внутренним кодами Рида — Соломона и образуют блоки кода, исправляющего ошибки (ECC-преобразование) [8].

4. Таблица записи (Recording Frame) — блок кода, исправляющего ошибки, представленного 16 подблоками, каждый из который подается на вход коду с ограниченной длиной поля записи (Run-length Limited, RLL).

5. Физический сектор (Physical Sector) — создание физических секторов из таблиц записи. На данном шаге происходит добавление синхронизирующих последовательностей.

Данные после таблицы данных и таблицы скремблирования подаются на шаг конфигурации блоков кода, исправляющего ошибки (ECC Block configuration) [7].

Все данные с предыдущего шага (таблицы скремблирования) делятся на блоки размером 192×172 байт (рис. 4).

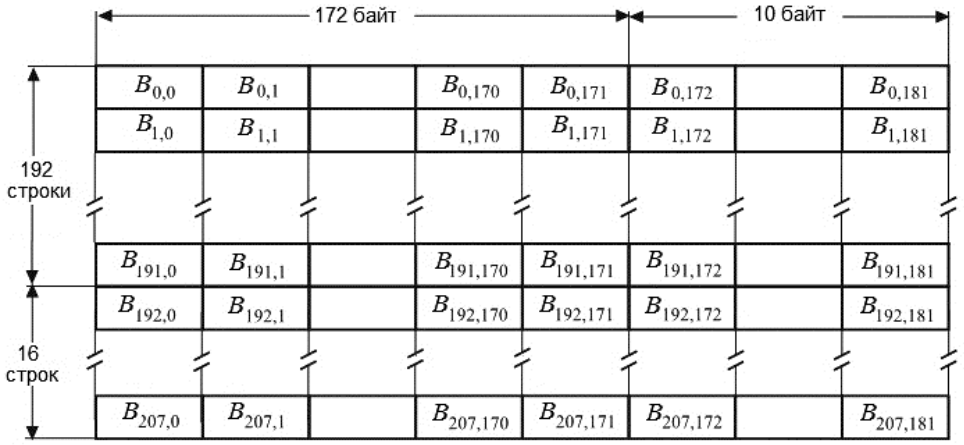

Рис. 4. Блоки размером 192×172 байт

Матрица размером 192×172 байт — информационная матрица. В этой матрице выбирается 172 столбца и каждый из них передается на вход как информационный вектор кода Рида — Соломона RS(208,192,17). На выходе кода — промежуточная матрица размером 208×172 байт. В полученной промежуточной матрице берется каждая из 208 строк и поступает как информационный вектор на вход кода RS(182,17,10). Полученная матрица 208 ×182 байт является кодовой матрицей. Первое преобразование называют внешним кодом, а второе — внутренним. Внешний код предназначен для исправления пакетных ошибок (возникают при дефектах производства диска, от длительного хранения, попадания пыли и проявления царапин на диске), а внутренний — для исправления одиночных ошибок (проявляются при сбое считывающей головки, а также при сбоях в цепях частей печатной платы) [9].

Данные стегосообщения можно записать в биты данных кодовой матрицы. Объем записываемых данных должен быть в пределах определенной нормы, зависящей от распределения ошибок в канале (канал представляет собой три канала: канал записи на диск, канал хранения диска, канал чтения с диска), для высокой вероятности декодирования кодовой матрицы. При отсутствии ошибок в канале, этот объем составит 182 × 16 = 2 916 байт на один блок кода, исправляющего ошибки, что составляет примерно 7,7 %.

Следует отметить, что ECC-преобразование не является конечным шагом преобразования информации перед записью на диск. После шага ECC, идет шаг создания таблиц записи, одним из подшагов которого является кодирование с помощью кодов с ограниченной длиной поля записи RLL(2,10) [7].

На RLL уровне каждые 8 бит преобразуются в последовательность длиной 16 бит (рис. 5). Таблицы преобразований описаны в работе [7]. Если полагать, что третья сторона имеет доступ к данным на шаге RLL, то она знает, какие ошибки являются настоящими, а какие фиктивными, т. е. содержащими сообщение. Чтобы избежать этого, необходимо записывать сообщение на уровне RLL.

Рис. 5. Схема кодирования с помощью кодов с ограниченной длиной поля записи

Стеганография в кодах исправляющих ошибки в дисках Blu-Ray. В стандарте [10] описан алгоритм кода, исправляющего ошибки в дисках Blu-Ray.

В отличие от DVD-дисков в кодах, исправляющих ошибки дисков Blu-Ray, нет внутреннего кодирования, предназначенного для исправления одиночных ошибок. Видимо с развитием техники сбои в цепях печатной платы оптического привода и в лазерной головке стали происходить настолько редко, что отпала необходимость кода, который исправляет одиночные ошибки.

Данные разбиваются на восемь блоков размером 216×38 байт — информационные матрицы (рис. 6). Затем каждая матрица подается на вход кода Рида — Соломона RS(248, 216, 32). Итоговый блок кода, исправляющего ошибки, также содержит четыре колонки с так называемыми символами застав (pickets), которые кодируются независимо кодом RS(9, 5, 4). Заставы применяются для обнаружения начала и конца пакетной ошибки. В отличие от DVD-дисков в дисках Blu-Ray избыточность составляет около 12,9 %.

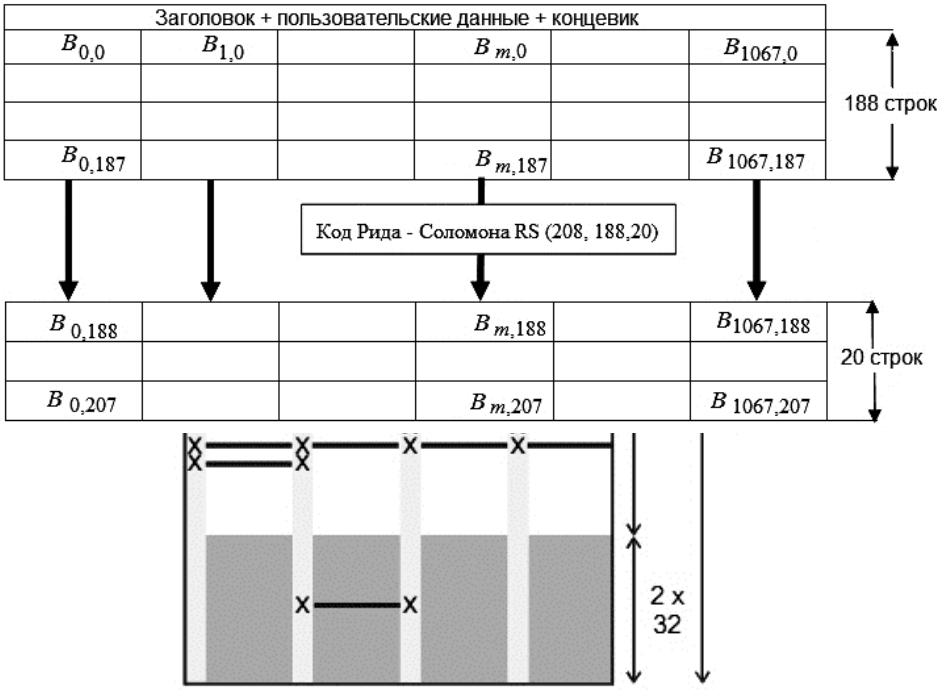

Рис. 6. Информационная матрица:
— информационные байты; — заставы; — проверочные байты; × — ошибки в заставах; —— — ошибки в коде, исправляющем ошибки

Стеганография в кодах, исправляющих ошибки в голографических многоцелевых дисках (Holographic Versatile Disc, HVD). Возможно скоро диски Blu-Ray заменят на голографические многоцелевые диски емкостью 300 Гбайт …1,6 Тбайт. В настоящее время компания InPhase Technologies уже продает первые серийные голографические диски емкостью 300 Гбайт.

Для таких дисков нет единого общего стандарта, поскольку идет активное развитие этих дисков [11, 12].

В HVD-дисках, как и в DVD-дисках применяются внешние и внутренние коды [12] Внешний код предназначен для исправления пакетных ошибок (рис. 7).

Из 188 строк данных образуется информационная матрица размером 1068×188 байт. Каждый столбец матрицы по 188 байт подается на вход кода RS(208, 188, 20). Избыточность составляет 9,62 %, что выше, чем у DVD-дисков, но меньше чем у дисков Blu-Ray. Тем не менее это пока ничего не значит, так как без знания о распределении ошибок на оптических дисках, невозможно сказать сколько именно байт в блоках кодов, исправляющих ошибки, можно использовать для стеганографии.

Требования к созданию алгоритма стеганографии в кодах, исправляющих ошибки в оптических дисках. Согласно изложен-

Рис. 7. Схема внешнего кода

ному выше, перед построением конкретных моделей стеганографии необходим статический анализ ошибок в кодах. После проведения расчета и анализа статистики при построении модели требуется, чтобы искомые данные контейнера имели ошибок не больше, чем может исправить декодер, основанный на кодах Рида — Соломона. Для обеспечения стойкости диск со скрытым сообщением должен иметь то же распределение ошибок, что и диск без скрытого сообщения, т. е. был бы идеальной стегосистемой. Для этого следует разработать средства и методы записи на диск (и возможно методы производства самого диска), отличающиеся от стандартных методов и средств.

Выводы. На основе изучения специальной литературы можно констатировать, что в настоящее время математический аппарат для анализа стегосистем недостаточно развит. Исследования в области стеганографии в кодах, исправляющих ошибки, весьма актуальны, так как в ближайшем будущем носители информации, использующие коды будут интенсивно развиваться и широко использоваться.

Запись и чтение с диска (DVD, Blu-Ray, HVD) имеют достаточно низкую скорость, что служит большим преимуществом стеганографии в кодах, исправляющих ошибки. Это связано с тем, что третья сторона будет тратить существенное время для извлечения контейнера, подозреваемого на наличие сообщения.

В стегоалгоритмах с кодами, исправляющими ошибки, контейнер может представлять собой произвольный набор символов, в отличие, например, от стегосистем с потерей информации, где контейнером может быть только изображение или аудиофайл. Это делает стегоалгоритмы с кодами еще более привлекательными с точки зрения практических приложений.

СПИСОК ЛИТЕРАТУРЫ

1. Pfitzmann B. Information Hiding Terminology. Results of an Informal Plenary Meeting and Additional Proposals // Springer Lecture Notes in Computer Science. 1996. Vol. 1174. — P. 347—350.

2. http://gumilev.ouc.ru/akrostih.html — сайт поэзии Николая Гумилева.
3. Wayner P. Mimic Functions // Cryptologia. 1992. XVI. № 3. — P. 193—214.
4. Барсуков В.С., Романцов А.П. Компьютерная стеганография, вчера, сегодня, завтра. Технологии информационной безопасности XXI века // Специальная техника. 1998. № 4—5.
5. Грибунин В.Г., Оков И.Н., Туринцев И.В. Цифровая Стеганография. — М.: Солон-Пресс, 2002. — С. 62—66.
6. Уточка Р.А., Фадин А.А., Шакалов И.Ю. Проблемные вопросы гарантированного уничтожения информации на носителях с полупроводниковой энергонезависимой перезаписываемой памятью // Вестник МГТУ им. Н.Э. Баумана. Спецвыпуск «Технические средства и системы защиты информации». 2011. — С. 7—19.
7. 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. Section 4. Part 19. Section 4. Part 16. Annex G.
8. Сагалович Ю.Л. Введение в алгебраические коды. — М.: ИППИ РАН, 2010. — С. 302.
9. Standard Blu-ray Disk Format: 1.B Physical Format Specifications for BD-R // 5th Edition. October, 2010. — P. 31.
10. Standard Blu-ray Disk Format: 1.A Physical Format Specifications for BD-RE // 3rd Edition. October, 2010. — P. 36—37.
11. International Standard ECMA-378. Information Interchange on Read-Only Memory Holographic Versatile Disc (HVD-ROM). Capacity 100 Gbytes per disk. 1st Edition. May, 2007.
12. 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.

Статья поступила в редакцию 4.07.2012