Современные направления стеганографии

Н. П. Варновский, Е. А. Голубев, О. А. Логачёв

1. Введение

Стеганография представляет собой специфическую область человеческой деятельности, связанной с разработкой и анализом методов сокрытия факта передачи информации. Подобно криптографии, стеганография известна со времен античности. Но на этом аналогии, по крайне мере в контексте теоретических исследований, заканчиваются. За последние четверть века возникла и успешно развивается новая математическая дисциплина — криптология, или, что то же самое, математическая криптография, изучающая математические модели криптографических схем. Попытки создания математической стеганографии (которую, быть может, следует именовать также стеганологией) предпринимаются, но исследования здесь находятся лишь в зачаточном состоянии.

Такое положение дел обусловлено прежде всего сложностью возникающих в стеганографии задач. Всякая попытка построения математических моделей стеганографических систем сопряжена с необходимостью рассмотрения большого количества случаев и подслучаев, не допускающих простой и единообразной трактовки. Другими словами, внешняя среда, в которой должны функционировать стеганографические системы, имеет гораздо большее, по сравнению с внешней средой криптографических схем, количество степеней свободы.

Но, тем не менее, теоретические исследования в области стеганографии ведутся и основная цель настоящей работы состоит в изложении некоторых результатов этих исследований.

Заголовок статьи следует понимать как сокращение. Полное название должно было бы звучать примерно так: «Современное состояние математических исследований в стеганографии. Основные направления, модели и понятия».

2. Модели стеганографических каналов

Впервые в открытой литературе модель стеганографического канала была описана Симмонсом в работе, представленной на конференцию Crypto'83 [14]. Из-за существовавших в то время запретов на публикации по стеганографии Симмонс сформулировал эту модель на языке задачи о двух заключенных.

Двое заключенных, Алиса и Боб, находящиеся в различных тюремных камерах, могут обмениваться посланиями. Но вся их переписка проходит через руки тюремного надзирателя Уэнди. Алиса и Боб должны выработать план побега, обмениваясь внешне безобидными текстами или картинками, не вызывающими подозрения у противника (Уэнди).

В соответствии с терминологией, согласованной на конференции Information Hiding: First International Workshop [13], те сообщения, которые Алиса и Боб пытаются передать друг другу втайне от противника, называются *встроенными (embedded) сообщениями*. Предлагается также обобщенный термин *встроенное сообщение* ⟨*тип данных*⟩, где ⟨*тип данных*⟩ заменяется, как и всюду ниже, соответствующим типом данных: текст, изображение (image) и т. п. Однако, по нашим наблюдениям, этот термин не прижился и вместо него обычно используется термин *скрытое (hidden) сообщение*. Тип данных обычно ясен из контекста и в большинстве случаев может быть опущен.

В стеганографической литературе зачастую предполагается, что сообщение, которое Алиса пытается передать Бобу, предварительно шифруется. В этом случае под скрытым сообщением понимается криптограмма.

Та информация, в которую встраивается скрытое сообщение, называется *контейнером (cover, carriage)*. Обобщенный термин — *контейнер* ⟨*тип данных*⟩. Соответственно, различают текстовые контейнеры, аудиоконтейнеры, видеоконтейнеры и т. п.

Если у Алисы и Боба имеется некоторая общая секретная информация, то последняя называется *секретным стегоключом* или просто *секретным ключом*. Заметим, что наличие у Алисы и Боба общего секретного ключа не является необходимым требованием для создания стеганографического канала. Существует стеганография с открытым ключом [2, 5] и даже так называемая чистая (безключевая) стеганография [5]. В настоящей статье эти два варианта не рассматриваются.

Для контейнера, содержащего встроенное в него скрытое сообщение, используется термин *стего* ⟨*тип данных*⟩ или просто *стего*. Контейнер, не содержащий скрытого сообщения, будем называть пустым контейнером. Отметим, что многие авторы используют термин стего для того контейнера, который Алиса передает Бобу, вне зависимости от того, является ли он и в самом деле стего или пустым контейнером. Чтобы избежать противоречия, можно принять соглашение, в соответствии с которым пустой контейнер — это стего с пустым скрытым сообщением.

Совокупность тех средств, которые Алиса и Боб используют для создания стеганографического канала, называется *стеганографической системой* или просто *стеганосистемой*. Здесь сказывается влияние криптографии и принятой в ней традиции выделять криптосистемы в отдельный тип. Правильнее было бы говорить о стеганографических протоколах.

Стеганосистема определяется прежде всего теми преобразованиями, Emb и Ext, которые Алиса и Боб используют соответственно для встраивания скрытого сообщения в контейнер и для извлечения этого сообщения из стего. Более формально, Алиса вычисляет

$$\text{стего} = \text{Emb}(\text{контейнер}, \text{скрытое сообщение}, \text{ключ}),$$

а Боб —

$$\text{скрытое сообщение} = \text{Ext}(\text{стего}, \text{ключ}).$$

При этом требуется, чтобы при любом данном ключе Боб с достаточно большой вероятностью извлекал из стего в точности то скрытое сообщение, которое было встроено в него Алисой.

Преобразование Emb является, вообще говоря, не всюду определенным. Для простоты изложения доопределим его, полагая для всякой тройки (контейнер, скрытое сообщение, ключ), не входящей в область определения, значение преобразования Emb равным пустой строке λ.

Существенным элементом описания стеганографической системы является предположение об источнике контейнеров. Возможны следующие три варианта:

- контейнер поступает извне. Это означает, что Алиса никак не влияет на его выбор. Своего рода предельный случай возникает, когда контейнер выбирается противником;
- Алиса выбирает контейнер из некоторого множества контейнеров. На практике, однако, это множество не имеет обычно сколько-нибудь точного описания, что и составляет одну из основных проблем построения математической теории стеганографических систем;
- контейнер генерируется стеганосистемой в процессе вычисления преобразования Emb. Фактически это означает, что контейнер здесь вообще ни при чем, а стеганосистема по скрытому сообщению и ключу сразу создает стего.

Легко понять, что эти предположения об источнике контейнеров являются, по существу, предположениями о противнике. Основное предположение о противнике состоит в том, что он будет пропускать по каналу связи те контейнеры, которые покажутся ему естественными. ***Проблема создания математической модели естественного контейнера является важнейшей исследовательской проблемой в стеганографии.*** На данный момент не известно даже, возможно ли в принципе создание такой модели.

В литературе по стеганографии упоминаются следующие три типа противников:

- *пассивный*. Уэнди действует так, как это описано в классической модели Симмонса;
- *активный*. Уэнди может вносить небольшие изменения в передаваемый контейнер. Как правило, предполагается, что ее вмешательство должно оставаться незаметным для Боба;
- *злоумышленный*. На действия Уэнди не накладывается никаких ограничений. Она может как угодно изменять контейнеры или даже вообще блокировать канал связи между Алисой и Бобом. Разумеется, от злоумышленного противника невозможно защититься. Но в реальности обычно имеется множество внешних обстоятельств, препятствующих злоумышленному поведению Уэнди.

Отсутствие четких критериев, отграничивающих активных противников от злоумышленных, — еще одна серьезная проблема, возникающая при построении математической теории стеганографических систем.

Противник может считаться активным также и в том случае, когда он проводит активную атаку на стеганосистему: атаку с выбором скрытого сообщения, атаку с выбором контейнера и т. п. (см. раздел 3).

Помимо термина «стеганографический канал» в литературе встречаются еще два: *скрытый (covert)* и *подпороговый (subliminal)* каналы. Поскольку в стеганографии еще нет общепринятой терминологии, возможно, что разные авторы вкладывают различный смысл в эти термины. Ниже мы приводим трактовки, которые нам удалось синтезировать на основе анализа работ по стеганографии.

Скрытый канал — это стеганографический канал, все внешнее поведение которого полностью описывается контейнерами. Иными словами, множество всех допустимых стего совпадает со множеством всех пустых контейнеров. Формально, стеганосистема для скрытого канала должна удовлетворять следующему требованию. Пусть C — множество всех контейнеров, K — множество ключей, M — множество скрытых сообщений и $S = \{s = \mathrm{Emb}(c, m, k),\ s \neq \lambda \,|\, c \in C,\ m \in M,\ k \in K\}$ — множество стего. Тогда $S \subseteq C$. Стеганографические каналы, создаваемые в операционных системах и других системах программного обеспечения, а также в криптографических протоколах, во многих случаях могут быть только скрытыми. Те из стеганографических каналов, которые не являются скрытыми, мы в дальнейшем будем называть общими.

Термин *подпороговый канал* был введен Симмонсом в связи с исследованием возможности создания скрытых каналов в криптографических протоколах (аутентификации, электронной подписи). По-видимому, он хотел подчеркнуть следующую особенность таких каналов. В обычном стеганографическом канале скрытое сообщение можно запрятать в один из «углов» контейнера в надежде, что противник не станет «шарить» по всем углам. Но если он посмотрит в нужный угол, то с большой вероятностью обнаружит скрытое сообщение. Криптографические протоколы позволяют строить скрытые каналы, существование которых, в предположении стойкости (криптографической) самого протокола, невозможно обнаружить в принципе. Для таких скрытых каналов целесообразно ввести специальный термин — подпороговый канал. В качестве примера можно рассмотреть вырожденный случай, когда Уэнди разрешает пересылку любых случайных последовательностей битов. Тогда шифр Вернама обеспечивает абсолютно необнаружимый скрытый канал. Термин «подпороговый» (subliminal) заимствован из психологии и означает, в развернутой формулировке, «находящийся ниже границы восприятия».

Возможно также, что единственной причиной появления словосочетания «подпороговый канал» послужила необходимость подыскать синоним, который мог бы заменить термин «стеганографический канал» в открытых публикациях.

При построении стеганосистем для скрытых каналов у стеганографа остается меньше свободы, поэтому данная задача в определенном смысле сложнее задачи построения стеганосистем для общего стеганографического канала. Но в то же время и возможности активного противника оказываются сильно ограниченными. Например, если речь идет о подпороговом канале в протоколе аутентификации, то активный противник может вносить в пересылаемые данные только такие изменения, которые не препятствуют выполнению протокола аутентификации в соответствии с его криптографическим определением.

Одной из важнейших характеристик стеганографического канала является его пропускная способность, под которой понимается максимальная длина скрытого сообщения, пересылаемого в одном контейнере. Существует вполне очевидное с интуитивной точки зрения, но не поддающееся формализации, разделение стеганографических каналов на *широкополосные* и *узкополосные*. Все возникающие в стеганографии научно-технические проблемы относятся к случаю широкополосных каналов. Если один или несколько битов скрытого сообщения передаются в огромном потоке, скажем аудио- или видеоинформации, то ясно, что такой стеганографический канал будет практически необнаружим. Но здесь возникает терминологическая проблема: где проходит граница между стеганосистемами для узкополосных каналов и системами передачи условных знаков. Например, Алиса должна передать Бобу всего один бит информации. Они заранее договорились, что Алиса будет пересылать сводку погоды, и если в этой сводке будет слово «облачно», то значение бита равно 0, а если слово «пасмурно», то оно равно 1. Являются ли системы передачи условных знаков предметом стеганографии? По этому поводу нет единого мнения даже у авторов настоящей статьи.

Так же как и в случае криптосистем, стойкость стеганосистемы (стеганостойкость) может быть

определена только относительно конкретной пары (атака, угроза). В разделе 3 рассматриваются угрозы безопасности стеганографических систем и обсуждается классификация атак на стеганосистемы.

Здесь уместно сделать два важных замечания.

1. По аналогии с подходом, принятым в криптографии, авторы теоретических работ по стеганографии предполагают, что стеганосистема известна противнику полностью, за исключением секретного ключа. Как и в криптографии, имеются два основных довода в пользу принятия такого предположения. Во-первых, без него не ясен сам предмет исследования. Во-вторых, всегда разумно обеспечить некоторый запас прочности, сделав предположение в пользу противника. Но все же для стеганографии это предположение представляется более далеким от реальности, чем для криптографии.

2. Говоря о противнике, авторы работ по стеганографии не уточняют, о каком противнике идет речь. Возможны, по крайней мере, два принципиально различных варианта. В первом из них Уэнди осуществляет выборочный беглый просмотр большого объема информации, пересылаемой, например, в компьютерной сети большим количеством абонентов. Во втором Уэнди прослушивает один канал связи между Алисой и Бобом и может потратить значительные усилия на выявление в нем стеганографического канала. Даже весьма поверхностного анализа литературы достаточно, чтобы понять, что авторы работ, посвященных синтезу стеганографических систем, обычно подразумевают первый вариант, а авторы работ по их анализу — второй.

Специалисты, занимающиеся анализом стеганосистем с целью обнаружения слабостей в их конструкциях, называются *стеганалитиками*, а область их деятельности — *стеганализом*.

3. Атаки на стеганографические системы и угрозы их безопасности

Ниже мы перечисляем основные угрозы безопасности стеганографических систем.

Обнаружение стеганографического канала. Это — самая слабая из угроз безопасности стеганосистем. Она может быть осуществлена пассивным противником. Сама семантика слова «стеганография» требует признать стеганосистему нестойкой, если противник может обнаружить создаваемый ею стеганографический канал. Поэтому о защите от этой угрозы часто говорят как об основной задаче стеганографии. Заметим, однако, что в большинстве работ не уточняется, что понимается под обнаружением стеганографического канала. Здесь имеются две возможности:

- передается последовательность контейнеров и либо все они пустые, либо часть из них являются стего, созданными с помощью одной и той же стеганосистемы (такой стеганографический канал называется каналом с повторениями). Если не все контейнеры пустые, то противник должен рано или поздно установить этот факт;
- передается один контейнер и противник должен распознать, содержит ли этот контейнер встроенное сообщение (канал без повторений).

Возможен и такой сценарий. Уэнди получила некоторую последовательность контейнеров, которые Алиса пересылала Бобу. В результате анализа этих контейнеров Уэнди могла установить наличие стеганографического канала и даже полностью или частично прочитать скрытые сообщения. После этого Алиса передает еще один контейнер. Требуется выяснить, содержит ли он скрытое сообщение. Но подобные вариации определяются уже не столько угрозой безопасности стеганосистемы, сколько типом атаки на нее (см. ниже).

Подчеркнем, что во всех вариантах задача противника состоит в том, чтобы отличить стего от пустого контейнера.

Извлечение скрытого сообщения. Противник должен найти скрытое сообщение, содержащееся в данном стего. Существует и более слабый вариант этой угрозы: противник должен получить какую-либо частичную информацию о скрытом сообщении. Эта угроза также может быть осуществлена пассивным противником.

Разрушение скрытого сообщения. Такая угроза существует только со стороны активного противника. В этом случае Уэнди должна внести в контейнер такие допустимые изменения, чтобы в результате Боб не смог извлечь из него скрытое сообщение (если таковое в нем было).

Подмена скрытого сообщения. Это — самая сильная из угроз; она может быть осуществлена только активным противником. Содержащееся в контейнере скрытое сообщение Уэнди должна заменить другим, выгодным ей, сообщением так, чтобы Боб не заподозрил подмену.

Существуют две основные характеристики угроз безопасности стеганографических систем. Сила угрозы определяется тем, насколько серьезными могут быть последствия ее осуществления для Алисы и Боба. Как уже отмечалось выше, подмена скрытого сообщения — самая сильная из угроз. Что же касается сложности осуществления угроз, то здесь ситуация не настолько простая, как это может показаться на первый взгляд. В литературе нередко можно встретить утверждения, что активный противник имеет значительное преимущество в том смысле, что разрушить скрытое сообщение значительно проще, чем обнаружить стеганографический канал. Например, если скрытое сообщение пересылается в младших битах пикселей изображения, то активному противнику достаточно заменять эти биты во всех контейнерах (независимо от того, являются ли они пустыми) на случайные. Однако, существуют сценарии (см. подраздел 5.3), в которых активный противник не имеет существенных преимуществ.

Известны следующие основные типы атак на стеганографические системы (см. [11]).

Атака с известным стего. Самая слабая из всех возможных атак, которую всегда может провести пассивный противник. В случае стеганографического канала без повторений предполагается, что в распоряжении Уэнди имеется только контейнер, который Алиса передавала Бобу. На основе анализа этого контейнера Уэнди должна осуществить свою угрозу безопасности стеганосистемы (обнаружить, извлечь, разрушить, подменить). Для стеганографического канала с повторениями возможны два варианта этой атаки:

1. Уэнди получает некоторую последовательность контейнеров, пересылаемых Алисой Бобу. Если среди этих контейнеров имеются стего, то предполагается, что все они созданы с помощью одной и той же стеганосистемы. Если угрозой является обнаружение стеганографического канала, то данная атака очевидным образом сводится к предыдущему случаю (одного контейнера). Для остальных типов угроз ситуация несколько сложнее, но также позволяет перейти, при подходящей переформулировке угрозы, к случаю одного контейнера.

2. Уэнди получает последовательность $C = \{c_1, \ldots, c_n\}$ контейнеров, пересылаемых Алисой Бобу. Кроме того, Уэнди известно, что все контейнеры из некоторого подмножества $C_1 \subseteq C$ пустые, а контейнеры из $C_2 \subseteq C$ являются стего. Как и прежде предполагается, что все стего из C, а также контейнер c_{n+1}, если он не пустой (см. ниже), созданы одной и той же стеганосистемой. Уэнди получает также информацию, частичную или полную, о скрытых сообщениях, которые содержатся в стего из множества $C_3 \subseteq C_2$. Далее Уэнди получает контейнер c_{n+1} и должна на основе его анализа осуществить угрозу безопасности стеганосистемы. Принципиальное отличие от случая 1 в том, что угроза безопасности, какой бы она ни была, относится только к контейнеру c_{n+1}.

Атака с известным контейнером. Уэнди получает контейнер, пересылаемый Алисой Бобу, и соответствующий ему пустой контейнер. Детерминированный случай тривиален. Более содержателен следующий сценарий. В каждый контейнер перед передачей по каналу связи между Алисой и Бобом вносятся небольшие случайные искажения. Уэнди знает исходный пустой контейнер и контейнер, передаваемый по каналу связи. Ее основная задача — понять, что содержится в последнем — случайный шум или скрытое сообщение.

Естественным усилением данной атаки является атака с выбором контейнера, когда Уэнди имеет возможность выбрать исходный пустой контейнер.

Атака с известным скрытым сообщением. Эта атака возможна только в случае стеганографического канала с повторениями и может быть осуществлена пассивным противником. Предполагается, что Уэнди знает стего и, быть может, соответствующий пустой контейнер. Кроме того, она каким-то образом узнает скрытое сообщение, встроенное в стего, и использует эту информацию для анализа используемой стеганосистемы (например, пытается определить секретный ключ, чтобы реализовать какие-либо угрозы безопасности стеганосистемы в будущем).

Атака с выбором скрытого сообщения. Аналогична предыдущей, но противник может сам выбирать скрытое сообщение. Разумеется, такую атаку может провести только активный противник. Возможен, например, следующий сценарий. Уэнди «подбрасывает» нужное ей сообщение Алисе и, получив стего, пытается определить секретный ключ стеганосистемы.

4. Теоретико-информационный подход

В настоящем разделе рассматривается теоретико-информационный подход к определению стойкости стеганосистем для стеганографических каналов без повторений в присутствии пассивного противника, обладающего неограниченными вычислительными возможностями.

В основе всех известных определений стойкости стеганосистем лежит требование неотличимости распределения вероятностей на множестве стего от распределения вероятностей на множестве пустых контейнеров. В данном разделе рассматривается статистическая неотличимость, или, иначе говоря, неотличимость относительно произвольных алгоритмов. В разделе 5 исследуется стойкость, определяемая вычислительной неразличимостью, т. е. неотличимостью относительно алгоритмов с ограничениями на вычислительные ресурсы.

Парадигма неотличимости распределений вероятностей заимствована из математической криптографии. Заметим, однако, что ее адекватность для стеганографии не очевидна. По крайней мере, в случае стеганографического канала без повторений не ясно, насколько оправданными будут усилия отправителя по имитации распределения вероятностей на множестве пустых контейнеров. Не следует ли вместо этого стремиться передать скрытое сообщение в одном из наиболее вероятных контейнеров?

Всюду далее в данной работе используется следующая терминология. У отправителя имеются контейнер, называемый исходным, а также функция (алгоритм) G, позволяющая преобразовать исходный контейнер в контейнер с шумом. Например, в случае контейнеров, содержащих визуальную информацию, существует некоторый объект, такой как картина, пейзаж и т. п. Исходный контейнер — это электронная фотокопия объекта. Алгоритм G позволяет вносить в эту фотокопию случайные искажения, соответствующие небольшим изменениям уровня освещенности, угла, под которым камера направлена на объект и т. п. Вполне очевидна идея создания стеганографического канала путем маскировки скрытого сообщения под шум, вносимый алгоритмом G в исходный контейнер. Предполагается, что Уэнди с большой вероятностью пропускает по каналу связи пустые контейнеры с шумом, создаваемые алгоритмом G. Поэтому последние называются также допустимыми контейнерами.

4.1. Модель стеганосистемы

Пусть $\mathcal{C}$, $\mathcal{K}$ и $\mathcal{M}$ — некоторые конечные множества. В рассматриваемой модели контейнеры (как исходный, так и контейнер с шумом), секретный ключ и скрытое сообщение являются случайными величинами на множествах $\mathcal{C}$, $\mathcal{K}$ и $\mathcal{M}$ соответственно. Мы будем обозначать через C, K и M случайные величины, являющиеся соответственно исходным контейнером, секретным ключом и скрытым сообщением.

Обозначим через R случайную величину на некотором конечном множестве $\mathcal{R}$. Эта случайная величина используется Алисой для внесения дополнительной случайности в стего.

В рассматриваемой модели Алиса и Боб имеют секретный ключ k, являющийся значением случайной величины K. Чтобы передать Бобу сообщение m (являющееся значением случайной величины M), Алиса вычисляет стего $s = \mathrm{Emb}(c, m, k, r)$, где c и r — значения случайных величин C и R соответственно, и посылает его Бобу. Без ограничения общности можно считать, что $s \in \mathcal{C}$. Получив стего s, Боб может восстановить m путем вычисления $\mathrm{Ext}(s, k)$. Алиса также может послать Бобу значение случайной величины $G(c, D)$, где D — случайная величина на некотором конечном множестве $\mathcal{D}$, а G — некоторая функция из $\mathcal{C} \times \mathcal{D}$ в $\mathcal{C}$. Таким образом, D играет роль шума.

Противник (Уэнди) в рассматриваемой модели представляет собой семейство функций $\{A_c | c \in \mathcal{C}\}$, где $A_c \colon \mathcal{C} \to \{0, 1\}$. Цель противника состоит в том, чтобы, зная значение C, отличить стего (значение $S = \mathrm{Emb}(C, M, K, R)$) от контейнера с шумом (значения $C' = G(C, D)$). Если c — известный противнику исходный контейнер, то равенство $A_c(x) = 1$ ($A_c(x) = 0$) означает, что, по мнению противника, контейнер $x \in \mathcal{C}$ является стего (соответственно, является пустым).

Таким образом, рассматривается стойкость стеганосистемы против угрозы обнаружения стеганографического канала на основе атаки с известным контейнером.

Вышеуказанный противник может делать ошибки двух типов, называемые в математической статистике ошибками первого и второго рода. *Ошибкой первого рода* считается ситуация, когда Уэнди приняла пустой контейнер за стего. Если же Уэнди посчитала стего пустым контейнером, то имеет место *ошибка второго рода*. Обозначим через α и β средние (по C) вероятности ошибок первого и второго рода соответственно.

4.2. Определения стойкости

Из литературы известны две попытки определить теоретико-информационную стойкость стеганосистем. Определение Кашена [4] основано на следующем требовании: энтропия пустого контейнера (контейнера с шумом) относительно стего должна быть мала. Подчеркнем, что речь идет об относительной (relative) энтропии; см. определение ниже. Таким образом, Кашен рассматривает задачу противника по различению пустого контейнера и стего как задачу проверки статистических гипотез. Другой подход описан в работе Цёлльнера и др. [15]. Он основан на таком требовании: знание контейнера и соответствующего ему стего не уменьшает энтропию скрытого сообщения. Заметим, что здесь под задачей противника по существу понимается извлечение некоторой информации о скрытом сообщении (очевидно, что обнаружение стеганографического канала является частным случаем этой задачи).

Известна также работа Андерсона и Птиколя [3] и ее ранняя версия [2], в которых формулируются некоторые математические утверждения (например, оценка емкости стеганографических каналов сверху через разность энтропий). Однако эти работы являются обзорными и не дают математически строгих определений рассматриваемых понятий.

Пусть X, Y и Z — случайные величины с конечными носителями $\mathcal{X}$, $\mathcal{Y}$ и $\mathcal{Z}$ соответственно. Через $\mathrm{Ent}(X)$ обозначается шенноновская энтропия случайной величины X. Пусть также $\mathrm{Ent}(X|Y)$ — условная энтропия X при условии Y. Хорошо известно, что $\mathrm{Ent}(X|Y) \leqslant \mathrm{Ent}(X)$, причем $\mathrm{Ent}(X|Y) = \mathrm{Ent}(X)$ тогда и только тогда, когда X и Y независимы. В качестве меры близости распределений случайных величин X и Y будет использоваться *энтропия X относительно Y*, обозначаемая через $\mathrm{REnt}(X\|Y)$ и определяемая формулой

$$\mathrm{REnt}(X\|Y) = \sum_{x\in\mathcal{X}} \Pr\{X = x\} \log_2 \frac{\Pr\{X = x\}}{\Pr\{Y = x\}}.$$

В литературе энтропия X относительно Y называется также дискриминацией между X и Y, информационной дивергенцией X и Y, дивергенцией Кульбака — Лейблера X и Y, расстоянием Кульбака. Заметим, что значение $\mathrm{REnt}(X\|Y)$ конечно тогда и только тогда, когда $\mathcal{X} \subseteq \mathcal{Y}$. В частности, если $\Pr\{X = x\}$ мала (но отлична от 0) и $\Pr\{Y = x\} = 0$ для некоторого $x \in X$, то $\mathrm{REnt}(X\|Y) = +\infty$, что не вполне соответствует интуитивному представлению о близости распределений. Определим *энтропию X относительно Y при условии Z* следующей формулой:

$$\mathrm{REnt}((X|Z)\|(Y|Z)) = \sum_{z\in\mathcal{Z}} \Pr\{Z = z\}\, \mathrm{REnt}((X|Z = z)\|(Y|Z = z)).$$

Здесь через $\mathrm{REnt}((X|Z = z)\|(Y|Z = z))$ обозначена энтропия X относительно Y при условии $Z = z$, определение которой можно получить, заменив в приведенной выше формуле для $\mathrm{REnt}(X\|Y)$ все вероятности на условные, при условии $Z = z$. Известно, что $\mathrm{REnt}(X\|Y) \geqslant 0$, причем $\mathrm{REnt}(X\|Y) = 0$ тогда и только тогда, когда X и Y распределены одинаково. Поэтому $\mathrm{REnt}((X|Z)\|(Y|Z)) \geqslant 0$, причем $\mathrm{REnt}((X|Z)\|(Y|Z)) = 0$ тогда и только тогда, когда $(X|Z = z)$ и $(Y|Z = z)$ распределены одинаково для любого $z \in \mathcal{Z}$.

Пусть ε — произвольное вещественное неотрицательное число.

Определение 1 ([4])**.** Стеганосистема $(\mathrm{Emb}, \mathrm{Ext})$ называется

- *ε-стойкой по Кашену*, если $\mathrm{REnt}(C'\|S) \leqslant \varepsilon$;
- *абсолютно стойкой по Кашену*, если она является 0-стойкой по Кашену, т. е. если случайные величины C' и S распределены одинаково.

В этом определении речь идет о стойкости стеганосистемы против угрозы обнаружения стеганографического канала на основе атаки с известным стего. Определение допускает следующее очевидное обобщение на случай атаки с известным контейнером.

Определение 2. Стеганосистема $(\mathrm{Emb}, \mathrm{Ext})$ называется

- *ε-стойкой (по Кашену) против угрозы обнаружения стеганографического канала на основе атаки с известным контейнером*, если $\mathrm{REnt}((C'|C)\|(S|C)) \leqslant \varepsilon$.

Определение 3 ([15]). Стеганосистема $(\mathrm{Emb}, \mathrm{Ext})$ называется

- *ε-стойкой по Цёлльнеру и др.*, если $\mathrm{Ent}(M) - \mathrm{Ent}(M|C,S) \leqslant \varepsilon$;
- *абсолютно стойкой по Цёлльнеру и др.*, если она является 0-стойкой по Цёлльнеру и др., т. е. если случайные величины M и (C,S) независимы.

4.3. Результаты

Пусть стеганосистема $(\mathrm{Emb}, \mathrm{Ext})$ является ε-стойкой по Кашену. Тогда для оценки средних вероятностей α и β ошибок первого и второго рода Кашен [4] предлагает следующий хорошо известный в математической статистике метод. Известно, что для любой функции f на множестве $\mathcal{X} \cup \mathcal{Y}$ справедливо неравенство.

$$\mathrm{REnt}(f(X)\|f(Y)) \leqslant \mathrm{REnt}(X\|Y). \tag{1}$$

Для произвольных вещественных ξ и η таких, что $0 \leqslant \xi, \eta \leqslant 1$, положим

$$\delta(\xi,\eta) = \begin{cases} \xi\log_2(\xi/(1-\eta)) + (1-\xi)\log_2((1-\xi)/\eta), & \text{если } \xi \notin \{0,1\}; \\ \log_2(1/\eta), & \text{если } \xi = 0; \\ \log_2(1/(1-\eta)), & \text{если } \xi = 1. \end{cases}$$

Тогда из неравенства (1), неравенства Йенсена и определения ε-стойкости по Кашену вытекает следующая теорема.

Теорема 1 ([4]). *Пусть* $(\mathrm{Emb}, \mathrm{Ext})$ *— стеганосистема, ε-стойкая по Кашену. Тогда*

$$\delta(\alpha,\beta) \leqslant \varepsilon.$$

В частности, если $\alpha = 0$, то $\beta \geqslant 2^{-\varepsilon}$.

Очевидно, что утверждение этой теоремы справедливо и для стеганосистемы, ε-стойкой против угрозы обнаружения стеганографического канала на основе атаки с известным контейнером (в смысле определения 2).

В следующей теореме рассматривается детерминированный случай, когда C' совпадает с исходным контейнером, т. е. когда функция G тождественна.

Теорема 2 ([15]). *Пусть* $(\mathrm{Emb}, \mathrm{Ext})$ *— стеганосистема, абсолютно стойкая по Цёлльнеру и др. Тогда*

$$\mathrm{Ent}(S|C) = \mathrm{Ent}(C|S) = 0.$$

Этот результат показывает, что в детерминированном случае абсолютная стойкость невозможна, и теоретически обосновывает необходимость недетерминированности функции G.

5. Теоретико-сложностной подход

Как известно, в математической криптографии всякое определение стойкости криптографического протокола может быть сформулировано в «алгоритмическом стиле». Отсутствие метода взлома криптографического протокола формализуется посредством требования несуществования алгоритма, решающего стоящую перед противником задачу. Если на алгоритмы, используемые противником, не накладывается никаких ограничений, то говорят об абсолютной (синонимы — шенноновская, теоретико-информационная) стойкости. Формализации, получаемые при ограничении вычислительных ресурсов противника, дают определение теоретико-сложностной стойкости.

Ситуация в математической стеганографии, в целом, аналогична. Существуют теоретико-информационный подход, обсуждавшийся в разделе 4, и теоретико-сложностной подход, которому посвящен настоящий раздел. Не следует, однако, забывать, что, как уже отмечалось во введении, внешняя среда, в которой должны функционировать стеганосистемы, имеет большое количество степеней свободы. Поэтому абсолютная стойкость стеганосистемы на самом деле весьма относительна. В частности, в большинстве случаев источником контейнеров служит достаточно сложный и не контролируемый

отправителем (Алисой) физический процесс. При разработке стеганосистемы используется математическая модель этого источника. И если стеганосистема является абсолютно стойкой относительно модельного источника контейнеров, то при переходе к реальному источнику эта стойкость может быть потеряна.

Такого же рода проблема возникает и в случае теоретико-сложностного подхода. Но есть и существенное различие. В самом деле, если противник создал, имея доступ к источнику контейнеров и используя только эффективные алгоритмы, достаточно точную модель этого источника, то почему бы разработчику стеганосистемы не сделать то же самое?

5.1. Модель стеганосистемы

Всюду ниже $\mathbb{N}$ обозначает множество натуральных чисел, $n \in \mathbb{N}$ — параметр безопасности. Функции из $\mathbb{N}$ в $\mathbb{N}$, имеющие порядок роста n^γ, где $\gamma = \text{const}$, $\gamma > 0$, мы будем называть полиномиально растущими.

Пусть $C = \{C_n\}$, где $C_n \subseteq \{0,1\}^q$, — множество всех исходных контейнеров. Здесь $q = q(n)$ — полиномиально растущая функция. На множествах C_n не задается никакого распределения вероятностей. Не требуется также, чтобы эти множества были в каком-либо смысле плотными в соответствующих множествах $\{0,1\}^q$. В частности, можно считать, что для всякого n множество C_n содержит в точности один контейнер.

Определение 4. Последовательность контейнеров $\{c_n\}$, $c_n \in C_n$, называется полиномиально генерируемой, если существует (детерминированный) алгоритм, который на входе 1^n работает за полиномиальное (от n) время и выдает c_n.

Предполагается, что каждый исходный контейнер перед передачей по каналу связи подвергается некоторым случайным искажениям. Допустимые искажения задаются алгоритмически. Формально, пусть G — полиномиальная вероятностная машина Тьюринга, которая для всякого $c \in C_n$ вычисляет $c' = G(c)$, $c' \in \{0,1\}^q$. Здесь c' — случайная величина. Для дальнейшего нам удобнее рассматривать этот алгоритм G как детерминированный. Такой переход осуществляется с помощью следующего стандартного приема.

Поскольку машина G работает за полиномиальное время, существует полиномиально растущая функция $l\colon \mathbb{N} \to \mathbb{N}$ такая, что на любом входном слове длины n машина G использует не более $l(n)$ случайных битов. Пусть $c \in C_n$ и $r \in \{0,1\}^{l(n)}$. Тогда $c' = G(c,r)$. Иными словами, мы можем рассматривать машину G как детерминированную, интерпретируя содержимое ее случайной ленты как дополнительное входное слово. В дальнейшем мы будем называть алгоритм G генератором допустимых контейнеров. Этот алгоритм считается общеизвестным.

Мы рассматриваем модель стеганографического канала без повторений. Противник в нашей модели пассивный и полиномиально ограниченный. Это означает, что в его распоряжении имеется полиномиальная вероятностная машина Тьюринга A, которая на входе $s \in \{0,1\}^q$ выдает либо 0, либо 1. Если $A(s) = 1$, то контейнер s считается допустимым, т. е. Уэнди пропускает его по каналу связи между Алисой и Бобом. Если $A(s) = 0$, то контейнер s считается подозрительным и не пропускается.

Пусть $\alpha\colon \mathbb{N} \to (0,1]$. Основное предположение о действиях и возможностях противника следующее. Пусть $c \in C_n$, $r \in_R \{0,1\}^l$ и $c' = G(c,r)$. Тогда

$$\Pr\{A(c') = 1\} \geqslant \alpha(n).$$

Вероятность здесь определяется случайным выбором строки r и случайными величинами, которые использует в своей работе машина A. Интуитивно, $\alpha(n)$ — это порог, ограничивающий «злоумышленность» противника. Заметим, что в работах по стеганографии обычно предполагается, что используется «абсолютно надежный» источник контейнеров, для которого $\alpha \equiv 1$.

Стойкость стеганосистемы определяется относительно пары (атака с известным стего, угроза обнаружения стеганографического канала). Это означает, что противнику известна стеганосистема, за исключением ключа, а также контейнер s, передаваемый по каналу связи. Угроза состоит в обнаружении стеганографического канала, т. е. задача противника — установить, что представляет собой строка s — пустой контейнер или стего.

Определение 5 ([1])**.** Стеганосистема $(\mathrm{Emb}, \mathrm{Ext})$ для данного генератора допустимых контейнеров G называется стойкой относительно пары (атака с известным стего, угроза обнаружения стеганографического канала), если для любой полиномиально генерируемой последовательности контейнеров $\{c_n\}$,

для любой последовательности сообщений $\{m_n\}$, $m_n \in \{0,1\}^{t(n)}$, семейства распределений вероятностей $\{G(c_n)\}$ и $\{\mathrm{Emb}(c_n, m_n, k)\}$, где $k \in_R \{0,1\}^n$, полиномиально неразличимы.

Напомним, что два семейства распределений вероятностей $\{D_n\}$ и $\{D'_n\}$ называются полиномиально неразличимыми, если для любой полиномиальной вероятностной машины Тьюринга B, для любого полинома p и для всех достаточно больших n

$$\left| \Pr_{x \in D_n} \{B(x) = 1\} - \Pr_{x \in D'_n} \{B(x) = 1\} \right| < 1/p(n).$$

В нашем случае распределения D_n и D'_n заданы на двоичных строках длины $q(n)$.

Замечание 1. Требование полиномиальной генерируемости последовательности контейнеров $\{c_n\}$ можно заменить, например, требованием общедоступности исходных контейнеров и формализовать его посредством оракула, к которому могут обращаться все рассматриваемые алгоритмы. Тогда приводимые ниже результаты будет верны в универсуме, релятивизированном к этому оракулу.

Очевидно, что стойкость стеганосистемы в смысле определения 5 слабее теоретико-информационной стойкости и является тривиальным следствием последней. Для криптосистем, обладающих теоретико-информационной стойкостью, хорошо известна нижняя оценка Шеннона: длина ключа должна быть не меньше длины открытого теста. Аналогичные оценки доказаны и для стеганосистем, обладающих теоретико-информационной стойкостью (см., например, [15]). Нас же интересует случай $t(n) \geqslant n + 1$. Всюду ниже предполагается, что это неравенство выполняется для всех (достаточно больших) n.

5.2. Необходимое условие существования стойких стеганосистем

Теорема 3 ([1]). *Если существует стеганосистема с детерминированным алгоритмом* Emb*, стойкая в смысле определения 5, то существуют односторонние функции.*

Для доказательства теоремы 3 нам потребуются следующие два определения.

Определение 6. Пусть $\{D_n\}$ — семейство распределений вероятностей, где D_n определено на $\{0,1\}^{\beta(n)}$. Семейство $\{D_n\}$ называется полиномиально конструируемым (polynomially samplable), если существует полиномиальная вероятностная машина Тьюринга Samp такая, что $\mathrm{Samp}(1^n) = D_n$.

Определение 7 ([10]). Пусть $g\colon \{0,1\}^n \to \{0,1\}^{\beta(n)}$ — функция, вычислимая за полиномиальное время. Функция g называется генератором дополнительной энтропии, если существуют полином p и полиномиально конструируемое семейство распределений $\{D_n\}$ такие, что

1. Семейства распределений $\{D_n\}$ и $\{g(x)\}$, $x \in_R \{0,1\}^n$, полиномиально неразличимы.
2. $\mathrm{Ent}(D_n) \geqslant \mathrm{Ent}(g(x)) + 1/p(n)$.

Здесь $\mathrm{Ent}(Y)$ обозначает шенноновскую энтропию случайной величины Y.

Доказательство теоремы 3. Общая схема доказательства такова. В предположении существования стеганосистемы, стойкой в смысле определения 5, доказывается, что существуют генераторы дополнительной энтропии. В работе Импальяццо, Левина и Луби [10] (см. также [12]) доказано, что если существуют генераторы дополнительной энтропии, то существуют и псевдослучайные генераторы. Хорошо известно [10], что существование последних эквивалентно существованию односторонних функций.

Пусть существует стеганосистема, стойкая в смысле определения 5, и пусть G — соответствующий генератор допустимых контейнеров.

Генератор дополнительной энтропии g строится следующим образом. Пусть $\{c_n\}$ — полиномиально генерируемая последовательность контейнеров. Зафиксируем какой-либо полиномиальный (детерминированный) алгоритм, который на входе 1^n выдает сообщение $m_n \in \{0,1\}^{t(n)}$. Пусть $x \in \{0,1\}^n$ — входное слово генератора g. Положим $g(x) = \mathrm{Emb}(c_n, m_n, x)$. Очевидно, что функция g вычислима за полиномиальное время. Очевидно также, что $\mathrm{Ent}(g(x)) \leqslant n$.

Далее, согласно определению 5 семейства распределений $\{g(x)\}$, $x \in_R \{0,1\}^n$, и $\{G(c_n)\}$ полиномиально неразличимы. Напомним, что каждое распределение вероятностей из семейства $\{G(c_n)\}$

определяется равновероятным выбором строки r из $\{0,1\}^l$. Легко видеть, что семейства распределений $\{G(c_n)\}$ и $\{\mathrm{Emb}(c_n, m_n, k)\}$, где $k \in_R \{0,1\}^n$, а $m_n \in_R \{0,1\}^{t(n)}$, также полиномиально неразличимы. Из этого следует полиномиальная неразличимость семейств $\{g(x)\}$, $x \in_R \{0,1\}^n$, и $\{\mathrm{Emb}(c_n, m_n, k)\}$, $k \in_R \{0,1\}^n$, $m_n \in_R \{0,1\}^{t(n)}$. Ясно также, что семейство $\{\mathrm{Emb}(c_n, m_n, k)\}$, $k \in_R \{0,1\}^n$, $m_n \in_R \{0,1\}^{t(n)}$, полиномиально конструируемо и его энтропия не меньше $t(n)$ (это следует из однозначности извлечения скрытого сообщения m_n из стего $\mathrm{Emb}(c_n, m_n, k)$ алгоритмом Ext). По предположению $t(n) \geqslant n+1$, что доставляет требуемое неравенство для энтропий. □

Замечание 2. Можно предположить, что существование односторонних функций является необходимым условием для существования стойких стеганосистем и в том случае, когда алгоритм Emb вероятностный. Вопрос о том, так ли это на самом деле, остается открытым.

На первый взгляд может показаться, что существует тривиальное доказательство теоремы 3 посредством следующего рассуждения: для стойких стеганосистем функция, вычисляемая алгоритмом Emb, должна быть односторонней. В противном случае алгоритм ее инвертирования будет по заданному стего с достаточно большой вероятностью находить пары (ключ, сообщение), что, казалось бы, можно использовать, чтобы отличать стего от пустых контейнеров. Однако, из условия теоремы и определения 5 следует лишь существование генератора допустимых контейнеров G, для которого имеется стойкая стеганосистема. О распределении вероятностей, создаваемом алгоритмом G на множестве контейнеров, ничего не известно. Поэтому неясно, как будет вести себя алгоритм инвертирования на входах, выбранных из этого распределения. Более того, в случае стеганосистемы для скрытого канала носители распределений $G(c_n)$ и $\mathrm{Emb}(c_n, m_n, k)$, $k \in_R \{0,1\}^n$, совпадают для всякого n.

Вопрос о том, является ли существование односторонних функций достаточным условием для существования стеганосистем, стойких в смысле определения 5, остается открытым. Несложно доказать, что это условие и в самом деле достаточно в случае, когда генератор допустимых контейнеров является частично обратимым.

Определение 8. Пусть G — генератор допустимых контейнеров и пусть для всякой строки $r \in \{0,1\}^l$, $r = r_1 \| r_2$, где $r_1 \in \{0,1\}^{l_1(n)}$, $l_1(n)$ — полиномиально растущая функция из $\mathbb{N}$ в $\mathbb{N}$ такая, что $l_1(n) < l(n)$. Генератор G называется частично обратимым, если существует полиномиальный алгоритм Der такой, что для всякого c и всякой строки $r \in \{0,1\}^l$, $\mathrm{Der}(G(c,r)) = r_1$.

Заметим, что хотя данный сценарий с теоретической точки зрения не слишком интересен, именно для него разработано большинство описанных в литературе конкретных стеганосистем.

Всюду в дальнейшем будем предполагать, что $l_1(n) \geqslant t(n)$.

В данном случае мы рассматриваем стойкость стеганосистемы относительно пары (атака с выбором контейнера и с выбором скрытого сообщения, угроза обнаружения стеганографического канала). Под противником понимается полиномиальная вероятностная машина Тьюринга A, которая удовлетворяет сформулированному выше требованию (ограничивающему ее поведение на входах, созданных генератором допустимых контейнеров). Атака на стеганосистему состоит из двух фаз. На первой фазе машина A получает на вход 1^n и выдает $m_n \in \{0,1\}^{t(n)}$ и $c_n \in \{0,1\}^{q(n)}$. Разумеется, в данном случае m_n и c_n — случайные величины.

На второй фазе машина A получает на вход строку $s \in \{0,1\}^q$ и выдает либо 0, либо 1. Пусть $P_1(n) = \Pr\{A(s) = 1\}$ для случая, когда $s = G(c_n, r)$. Эта вероятность определяется выбором случайной строки r и случайными величинами самой машины A. Пусть $P_2(n) = \Pr\{A(s) = 1\}$ в случае, когда $s = \mathrm{Emb}(c_n, m_n, k)$. Вероятность P_2 определяется случайным выбором секретного ключа k и случайной строкой машины A.

Предполагается, что противник знает стеганосистему, за исключением ключа, и, кроме того, ему известны алгоритмы G и Der.

Определение 9. Стеганосистема $(\mathrm{Emb}, \mathrm{Ext})$ для данного генератора допустимых контейнеров G называется стойкой относительно пары (атака с выбором контейнера и с выбором скрытого сообщения, угроза обнаружения стеганографического канала), если для любой полиномиальной вероятностной машины Тьюринга A указанного выше типа, для любого полинома p и для всех достаточно больших n

$$|P_1(n) - P_2(n)| < 1/p(n).$$

Утверждение следующей теоремы носит в определенном смысле условный характер, поскольку оно предполагает существование частично обратимых генераторов допустимых контейнеров. Очевидно, что частично обратимый генератор можно построить без привлечения каких-либо недоказанных гипотез. Но для того, чтобы создаваемые им контейнеры были допустимыми, требуется соответствующее предположение о действиях и возможностях противника.

Теорема 4 ([1]). *Если существуют односторонние функции, то существуют стеганосистемы, стойкие относительно пары (атака с выбором контейнера и с выбором скрытого сообщения, угроза обнаружения стеганографического канала).*

Доказательство. Пусть $g\colon \{0,1\}^n \to \{0,1\}^{t(n)}$ — псевдослучайный генератор. Его существование следует из существования односторонней функции [10, 12]. Стеганосистема $(\mathrm{Emb}, \mathrm{Ext})$ строится следующим образом. Алгоритм Emb на входе (c_n, m_n, k) сначала вычисляет $\sigma = g(k) \oplus m_n$. Затем он вызывает алгоритм G, подавая ему на вход c_n и $r = \sigma \parallel r'$, где $r' \in_R \{0,1\}^{l(n)-t(n)}$. Пусть $s = G(c_n, r)$. Строка s (стего) и является выходом алгоритма Emb.

Алгоритм Ext сначала вызывает алгоритм Der, подавая ему на вход строку s. Пусть σ' — префикс длины $t(n)$ строки $\mathrm{Der}(s)$. Тогда $m_n = \sigma' \oplus g(k)$.

Доказательство стойкости построенной стеганосистемы вытекает непосредственно из определения псевдослучайного генератора: никакой полиномиальный вероятностный алгоритм не может отличить псевдослучайную строку $g(k)$, $k \in_R \{0,1\}^n$, от случайной строки той же длины. □

5.3. О модели с активным противником

В некоторых работах по стеганографии утверждается, что задача создания стеганографического канала в присутствии активного противника является очень сложной и требует разработки методов, отличных от тех, которые позволяют защищаться от пассивного противника. Главный итог подобных рассуждений — весьма пессимистический прогноз, поскольку на данный момент, по существу, никаких специальных методов защиты от активного противника не предложено. Тем не менее, в данном подразделе мы выделяем один частный (но весьма важный с практической точки зрения) случай, в котором от активного противника можно защититься теми же методами, что и от пассивного.

Как правило, рассуждения об активном противнике укладываются в такую схему: поскольку скрытое сообщение «упрятывается» отправителем в некоторый псевдошум, добавляемый к контейнеру, активный противник всегда может это сообщение разрушить путем модификации псевдошума. Но все это справедливо лишь в том случае, когда при передаче сообщений между Алисой и Бобом возникает шум в канале. Если же сообщения (например, файлы) передаются через компьютерную сеть, то они доставляются получателю практически без искажений. В данном подразделе мы рассматриваем именно этот случай бесшумного канала.

Далее, в классической задаче о двух заключенных предполагается, что Уэнди осуществляет цензурирование сообщений, пересылаемых между Алисой и Бобом, на законных основаниях. На практике, однако, такая ситуация встречается крайне редко. Поэтому мы будем предполагать, что активный противник должен скрывать свои действия от отправителя и получателя.

Итак, рассматривается следующая модель. Пусть, как и прежде, $c \in C_n$ — исходный контейнер, известный Алисе. С помощью алгоритма G она создает пустой контейнер $c' = G(c)$. Напомним, что m обозначает скрытое сообщение, $(\mathrm{Emb}, \mathrm{Ext})$ — стеганосистему, а k — ее секретный ключ. Алиса передает Бобу либо пустой контейнер c', либо стего $s = \mathrm{Emb}(c, m, k)$. Уэнди перехватывает передаваемый контейнер $\tilde{c}$ и с помощью полиномиального вероятностного алгоритма подмены Sub создает контейнер $c'' = \mathrm{Sub}(\tilde{c})$, который и передается Бобу. При этом действия алгоритма подмены должны быть незаметны для Боба, что формализуется требованием полиномиальной неразличимости семейств распределений $\{G(c)\}$ и $\{\mathrm{Sub}(c')\}$, $c' = G(c)$. Подчеркнем, что это условие никак не ограничивает действия Уэнди в том случае, когда по каналу связи пересылается стего.

Описанная выше модель подсказывает следующую методику использования стеганосистемы. Алиса должна послать Бобу скрытое сообщение m. Прежде чем отправить соответствующее стего s, Алиса выбирает некоторое количество исходных контейнеров $c_1, \ldots, c_u$, создает с помощью алгоритма G соответствующие пустые контейнеры $c'_1, \ldots, c'_u$, которые мы будем называть пробными, и посылает их Бобу. Назначение этой предварительной фазы — обнаружить присутствие в канале активного противника. Если на каком-либо из пробных контейнеров Боб обнаружит вмешательство Уэнди, то он подает Алисе условный знак (любым возможным способом) и стего не передается.

Разумеется, задача становится тривиальной, если Боб знает в точности все передаваемые пробные контейнеры $c'_1, \ldots, c'_u$ или может их отправить обратно Алисе (минуя Уэнди!) на проверку.

Параметр u считается неизвестным Уэнди и, вообще говоря, Бобу и может выбираться, например, случайным образом из множества $V = \{1, \ldots, v\}$. Мы считаем, что Уэнди не знает значение v. Это соответствует следующей реальной ситуации: по каналу связи между Алисой и Бобом передается некоторый поток контейнеров, среди которых могут быть как пустые, так и стего. Уэнди не имеет никакой априорной информации, которая позволяла бы разбивать этот поток на подпоследовательности связанных между собой контейнеров.

Определение 10. Стеганосистема $(\mathrm{Emb}, \mathrm{Ext})$ для данного генератора допустимых контейнеров G называется ε-стойкой против активного противника (против угрозы разрушения скрытого сообщения на основе атаки с известным стего), если:

1. Она является стойкой против пассивного противника (см. определение 5).
2. Для любой полиномиальной вероятностной машины Тьюринга Sub справедливо следующее. Пусть s' — контейнер, полученный Бобом, когда Алиса передавала стего s, т. е. либо $s' = s$, либо $s' = \mathrm{Sub}(s)$. Тогда $\Pr\{\text{Алиса передала } s \ \&\ \mathrm{Ext}(s') \neq m\} \leqslant \varepsilon$.

Фактически, п. 2 определения означает, что с помощью пробных контейнеров Алиса и Боб либо обнаруживают присутствие в канале активного противника, либо убеждаются, что канал достаточно безопасен и Боб с высокой вероятностью получит скрытое сообщение m.

Общая идея построения стеганосистемы, которая, предположительно, была бы стойкой в смысле определения 10, достаточно очевидна. Алиса выбирает $u \in_R V$ и создает пробные контейнеры $c'_1, \ldots, c'_u$, которые содержат заранее условленные метки. Если Боб обнаружит, что хотя бы одна из этих меток изменена, то он выдает «сигнал тревоги». В противном случае, т. е. когда такого сигнала не поступило, в качестве $(u+1)$-го контейнера Алиса передает стего.

Если G — частично обратимый генератор допустимых контейнеров, то описанную выше идею можно реализовать следующим образом. Секретным ключом стеганосистемы является пара (k_1, k_2), где $k_1, k_2 \in \{0,1\}^n$. Пусть $g\colon \{0,1\}^n \to \{0,1\}^{v \cdot l_1(n)}$ — псевдослучайный генератор. Напомним, что l_1 — длина той части случайной строки, используемой генератором допустимых контейнеров G, которая может быть извлечена из контейнера с помощью алгоритма Der (см. подраздел 5.2). Алиса вычисляет $g(k_1)$ и делит полученную последовательность на v блоков $r_1, \ldots, r_v$, каждый длины $l_1(n)$. Затем она выбирает исходные контейнеры $c_1, \ldots, c_u$, вычисляет пробные контейнеры $c'_1 = G(c_1, r_1), \ldots, c'_u = G(c_u, r_u)$ и пересылает их Бобу. Последний с помощью алгоритма Der извлекает из пробных контейнеров строки $r_1, \ldots, r_u$ и сравнивает их с соответствующими блоками последовательности $g(k_1)$. Ясно, что любое вмешательство Уэнди, которое привело к изменению хотя бы одного из этих l_1 битов хотя бы в одном пробном контейнере, будет обнаружено. В противном случае, Алиса, используя секретный ключ k_2 создает, как это описано в предыдущем подразделе, стего s, содержащее скрытое сообщение m, и посылает s Бобу. Очевидно, что для разрушения скрытого сообщения Уэнди должна либо угадать значение u, либо уметь отличать стего от пустых контейнеров.

5.4. Случай неравномерно распределенного шума

Выше мы рассматривали модель стеганосистемы в предположении существования частично обратимого генератора допустимых контейнеров. Нетрудно видеть, что это предположение на самом деле содержит в себе две гипотезы:

- существует эффективный алгоритм (частичного) извлечения шума из контейнера;
- шум, извлеченный из контейнера, распределен равномерно.

Как с практической, так и с теоретической точек зрения представляет интерес проблема построения стойких стеганосистем в случае, когда выполняется только первая из этих гипотез, т. е. шум имеет распределение, отличное от равномерного. В данном подразделе мы рассматриваем эту проблему при следующих предположениях:

- Алиса имеет доступ к источнику контейнеров как к оракулу;

- контейнеры сами являются «шумом», т. е. на каждый запрос оракул возвращает битовую строку, выбранную из распределения, близкого к равномерному;
- параметры распределения вероятностей на множестве контейнеров Алисе не известны. Она знает лишь, что это распределение близко к равномерному;
- рассматривается модель стеганографического канала без повторений;
- все контейнеры, полученные Алисой от оракула, пересылаются по каналу связи Бобу. Для передачи скрытого сообщения Алиса может выбрать любой из этих контейнеров. Все вопросы о том, каким образом Боб сможет понять, какой из контейнеров является стего, чтобы корректно извлечь из него скрытое сообщение, игнорируются.

Таким образом, рассматриваемая модель основывается на весьма сильных предположениях в пользу отправителя (или, точнее, в пользу разработчика стеганосистемы). Но поскольку в итоге будет получен отрицательный результат, такие предположения этот результат лишь усиливают.

Пусть n — параметр безопасности, $u = u(n)$ — полиномиально растущая функция. Алиса получает от оракула контейнеры $c_1, \ldots, c_u$, каждый длины n, выбирает $i \in \{1, \ldots, u\}$ и заменяет контейнер c_i на стего $s = \mathrm{Emb}(c_i, m, k)$, где k — секретный ключ стеганосистемы, а m — скрытое сообщение. Можно сделать еще одно предположение в пользу отправителя и считать скрытое сообщение m случайной, относительно равномерного распределения, строкой длины $t < n$.

В данной модели стойкость стеганосистемы относительно пары (атака с известным стего, угроза обнаружения стеганографического канала) определяется требованием вычислительной неразличимости семейств распределений $\{c_1, \ldots, c_{i-1}, c_i, c_{i+1}, \ldots, c_u\}$ и $\{c_1, \ldots, c_{i-1}, s, c_{i+1}, \ldots, c_u\}$. Для всякого значения параметра n распределение вероятностей из первого семейства определяется случайным выбором контейнеров, а распределение вероятностей из второго семейства — еще и случайным выбором индекса i, ключа k, скрытого сообщения m, а также случайными величинами алгоритма Emb. Заметим, что выбор индекса i не обязательно должен быть случайным и может рассматриваться как часть алгоритма Emb.

Мы не приводим формального определения стойкости, поскольку оно является очевидной модификацией определения 5.

Определение 11. Пусть D — распределение вероятностей на множестве $\{0,1\}^n$ и пусть X — случайная величина, принимающая значения в $\{0,1\}^n$ в соответствии с распределением D. Тогда min-энтропией случайной величины X называется величина

$$\mathrm{Ent}_{\min}(X) = \min_{x \in \{0,1\}^n} (-\log \Pr\{X = x\}).$$

Вместо min-энтропии можно рассматривать близкое ей понятие энтропии Реньи.

Определение 12. Пусть X и Y — независимые одинаково распределенные случайные величины. Энтропией Реньи случайной величины X называется величина

$$\mathrm{Ent}_{\mathrm{Ren}}(X) = -\log \Pr\{X = Y\}.$$

Из теории вероятностей известно, что min-энтропия и энтропия Реньи связаны между собой следующими неравенствами

$$\mathrm{Ent}_{\mathrm{Ren}}(X)/2 \leqslant \mathrm{Ent}_{\min}(X) \leqslant \mathrm{Ent}_{\mathrm{Ren}}(X).$$

Т. е. эти энтропии совпадают с точностью до мультипликативной константы 2. Кроме того, нетрудно показать, что энтропия Реньи всегда не превосходит шенноновскую энтропию.

Определение 13. Распределение $X = (X_1, \ldots, X_u)$ над множеством $\{0,1\}^{nu}$ называется (n,h)-блоковым источником, если для всех $i = 1, \ldots, u$ и для каждого $z \in \{0,1\}^{nu-n}$, $\mathrm{Ent}_{\min}(X_i | X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_u = z) \geqslant h$.

Здесь $\mathrm{Ent}_{\min}(X_i | X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_u = z)$ — условная min-энтропия случайной величины X_i при условии $X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_u = z$.

Под источником энтропии понимается $(n, n-1/p(n))$-блоковый источник, где p — некоторый фиксированный полином. Заметим, что (n,n)-блоковый источник соответствует равномерному распределению вероятностей на множестве всех битовых строк длины nu. Так что строки, выбираемые из $(n, n-1/p(n))$-блокового источника, весьма незначительно отличаются от чисто случайных строк.

Будем говорить, что стеганосистема описанного выше вида является *тривиальной*, если вероятность, что $s \neq c_i$, пренебрежимо мала как функция от n.

Теорема 5. *Если для данного фиксированного полинома p существует стеганосистема, стойкая для любого $(n, n-1/p(n))$-блокового источника контейнеров, то эта стеганосистема является тривиальной.*

Эта теорема является простым следствием следующей леммы из работы [9]. В формулировке леммы Γ — класс всех $(n, n-1/p(n))$-блоковых источников с u блоками, $N = nu$. Значения параметров u, N' и M ограничены сверху полиномами от n.

Лемма 1. *Пусть функции $F\colon \{0,1\}^N \times \{0,1\}^{N'} \to \{0,1\}^M$ и $G\colon \{0,1\}^N \to \{0,1\}^M$ и распределение вероятностей Y на множестве $\{0,1\}^{N'}$ таковы, что для любого распределения вероятностей $X \in \Gamma$ семейства распределений $\{F(X,Y)\}$ и $\{G(X)\}$ полиномиально неразличимы. Тогда вероятность $\Pr\{F(x,y) \neq G(x)\}$ пренебрежимо мала как функция от n. Здесь $x \in_R \{0,1\}^N$, а y выбирается из множества $\{0,1\}^{N'}$ в соответствии с распределением Y.*

Замечание 3. Из данной леммы в работе [9] выведен ряд результатов о несуществовании различных криптографических протоколов и примитивов (шифрование, разделение секрета, доказательства с нулевым разглашением и т. п.) в модели со слабыми источниками случайности (т. е. в модели, где источник чисто случайных битов заменен источником энтропии). При всем сходстве этих результатов с теоремой 5 имеется и принципиальное отличие. Выбор источника случайности для криптографических конструкций находится в руках разработчика, так что отрицательные результаты лишь указывают на необходимость проведения дополнительных исследований с целью отбора подходящих источников. А для стеганосистемы источник контейнеров представляет собой часть исходных данных.

5.5. Стеганография с «черным ящиком»

В работе Дедича и др. [8] исследуется возможность построения стойких стеганосистем в случае, когда источником контейнеров является оракул, или, иначе говоря, «черный ящик». Отправитель может выдавать запросы оракулу и в ответ получать пустые контейнеры, выбираемые из некоторого источника энтропии. Существенное отличие от модели, рассматривавшейся в предыдущем подразделе состоит в том, что отправитель имеет возможность принимать решение, что делать с полученным от оракула контейнером: преобразовать в стего и отослать получателю, или игнорировать.

Данная модель интересна тем, что позволяет уйти от далекого от реальности предположения, согласно которому отправителю (а, значит, и разработчику стеганосистемы) известны все требуемые параметры источника контейнеров, в т. ч. распределение вероятностей на множестве пустых контейнеров.

Рассматривается стеганографический канал без повторений. Стойкость стеганосистемы против угрозы обнаружения стеганографического канала на основе атаки с выбором скрытого сообщения определяется в теоретико-сложностном стиле [8]. Исследуется среднее количество запросов к оракулу, выдаваемых стеганосистемой, как функция от min-энтропии источника контейнеров. Доказана экспоненциальная нижняя оценка количества запросов.

Переходим к формальному изложению основного результата работы [8].

Пусть Σ — конечный алфавит. Под каналом C понимается отображение, которое преобразует предысторию $H \in \Sigma^*$ в распределение вероятностей D_H на множестве Σ. Предыстория $H = s_1 s_2 \dots s_l$ называется легальной, если всякий ее символ может быть получен, исходя из предшествующих символов, т. е. $\Pr_{D_{s_1 \dots s_{i-1}}}(s_i) > 0$.

В стеганосистемах, рассматриваемых в работе [8], доступ отправителя к каналу C формализуется посредством оракула M, который на запрос H возвращает символ s, представляющий собой реализацию случайной величины с распределением вероятностей D_H.

Определение 14. Стеганосистемой в модели с черным ящиком называется пара $S = (\mathrm{Emb}, \mathrm{Ext})$ полиномиальных вероятностных алгоритмов таких, что:

1. Входными данными алгоритма Emb служат секретный ключ k, строка $m \in \{0,1\}^*$ (скрытое сообщение), и предыстория H. Кроме того, этот алгоритм имеет доступ к оракулу M. Выходом алгоритма Emb является строка символов $s_1 s_2 \dots s_l \in \Sigma^*$ (стеготекст).

2. Входными данными алгоритма Ext служат секретный ключ k, предыстория H и стеготекст $s_1 s_2 \dots s_l \in \Sigma^*$. Алгоритм выдает скрытое сообщение $m \in \{0,1\}^*$.

Здесь $k \in_R \{0,1\}^n$, где n — параметр безопасности.

Авторы предполагают, что длина l стеготекста зависит только от длины скрытого сообщения m, но не от его содержания. Под скоростью стеганографического канала понимается число $w > 0$ такое, что для передачи скрытого сообщения длины lw требуется в среднем l символов стеготекста.

Надежность стеганосистемы. Надежность стеганосистемы S с параметром безопасности n, для канала C и сообщений длины l определяется как

$$\mathrm{Rel}_S = \min \Pr\{\mathrm{Ext}(H, \mathrm{Emb}^M(H, m, k), k) = m\}.$$

Минимум берется по всем легальным предысториям H и всем $m \in \{0,1\}^l$. Вероятность определяется случайным выбором ключа $k \in \{0,1\}^n$ и случайными величинами алгоритмов Emb и Ext и оракула M.

Рассматривается пассивный противник, проводящий атаку с выбором скрытого сообщения. Цель такого противника — отличить стего, содержащее выбранное им (противником) скрытое сообщение от случайной выборки из канала C. Для формализации определяется оракул $O(H, \cdot)$, который на входе (H, m), где $m \in \{0,1\}^*$, H — предыстория, действует следующим образом. Сначала O вычисляет длину l стеготекста, который выдал бы алгоритм Emb на входе (H, m, k), затем последовательно выбирает случайные s_i, каждое из распределения $D_{H \| s_1 \dots s_{i-1}}$, и выдает $s_1 \dots s_l$.

Определение 15. Вероятностный алгоритм W называется (t, d, q, l)-пассивным противником для стеганосистемы S, если:

1. W работает за время t в среднем (здесь учитывается время, необходимое алгоритму Emb для ответов на запросы противника) и имеет длину описания d.

2. W имеет доступ к каналу C, формализуемый посредством оракула M.

3. Второй оракул, к которому имеет доступ алгоритм W, — это либо $\mathrm{Emb}^M(\cdot, \cdot, k)$, либо $O(\cdot, \cdot)$. W может выдать этому оракулу в среднем q запросов суммарной длины l битов.

4. Выходом алгоритма W является один бит.

Стойкость стеганосистемы. Успешностью противника W против стеганосистемы S с параметром безопасности n и каналом C называется величина

$$\mathrm{Adv}_S(W) = \left| \Pr\left\{ W^{M, \mathrm{Emb}^M(\cdot,\cdot,k)} = 1 \right\} - \Pr\left\{ W^{M, O(\cdot,\cdot)} = 1 \right\} \right|.$$

Вероятности определяются случайными величинами алгоритма W и оракулов, а первая из вероятностей — еще и случайным выбором ключа k из множества $\{0,1\}^n$.

Стойкость стеганосистемы для канала C определяется как функция

$$\mathrm{Sec}_S(t, d, q, l) = 1 - \max \mathrm{Adv}_S(W).$$

Максимум берется по всем (t, d, q, l)-пассивным противникам W.

Теорема 6. *Пусть S — стеганосистема со стойкостью $1 - \varepsilon$, надежностью $1 - \rho$ и скоростью w. Предположим, что противник имеет доступ к дополнительному оракулу, позволяющему проверять принадлежность символа s_i носителю распределения D_i. Тогда существует канал с min-энтропией h, для которого вероятность, что алгоритм* Emb *при отсылке случайного скрытого сообщения длины lw делает не более N запросов к оракулу M, не превосходит*

$$\left(\frac{Ne}{l2^w}\right)^l + \rho + \varepsilon R,$$

а, следовательно, среднее число запросов к оракулу на один символ стеготекста не меньше

$$\frac{2^w}{e}(1/2 - \rho - \varepsilon R),$$

где $R = 1/(1 - 2^h/|S|)$.

Этот результат, так же как и теорема 5, демонстрирует невозможность построения стойких стеганосистем для источников энтропии (в качестве источников контейнеров). Подчеркнем, что речь идет об «универсальных» стеганосистемах, которые должны быть пригодны для любого источника контейнеров, энтропия которого достаточно велика. Эти отрицательные результаты не исключают возможности существования специализированной стеганосистемы для каждого отдельного источника энтропии.

6. Электронные водяные знаки

Электронные водяные знаки и их специальная разновидность, электронные отпечатки пальцев, в последние годы стали весьма популярной темой исследований, привлекшей внимание специалистов из различных областей математики. Вместе с тем, этим исследованиям, на наш взгляд, не хватает самого главного, а именно, адекватной с точки зрения потенциальных приложений и математически корректной постановки задачи.

Начнем с замечания, что не существует единого мнения по поводу отнесения электронных водяных знаков к стеганографии. Как известно, стеганографические методы призваны скрывать сам факт передачи информации. В противоположность этому, наличие водяных знаков в какой-либо информации может быть общеизвестным и задача состоит в их защите от удаления и (или) изменения. Возможны два выхода из этой ситуации. Первый состоит в расширительном толковании термина «стеганография», включающем и электронные водяные знаки. Альтернативная точка зрения признает существование научной дисциплины под названием «сокрытие информации» (information hiding). Вся область исследований, связанных с проблематикой электронных водяных знаков, входит, наряду со стеганографией, в эту научную дисциплину.

Электронный водяной знак представляет собой некий аналог скрытого сообщения. Это — специальные данные, добавляемые в различного рода информацию, представленную в электронной форме, интеллектуальным собственником этой информации. Назначение электронного водяного знака — служить доказательством для третьих лиц, что данная информация является интеллектуальной собственностью того, кто внес в нее этот водяной знак. Электронные водяные знаки могут быть индивидуализированы. С этой целью продавец, поставляющий клиентам файл с какой-либо информацией, помещает в каждую копию этого файла водяной знак с данными, идентифицирующими клиента, которому данная копия была продана. В дальнейшем, в случае обнаружения пиратских копий файла, продавец сможет идентифицировать клиента, который эти копии распространяет.

Однако, индивидуализированные водяные знаки не позволяют продавцу доказывать третьим лицам, что пиратские копии распространяет именно данный клиент. Разновидность водяных знаков, обеспечивающая возможность такого доказательства, получила название электронных отпечатков пальцев. Доказательство может быть убедительным для третьих лиц только в том случае, если у каждого клиента имеется (сертифицированный) открытый ключ, а отпечаток пальцев содержит информацию, которую невозможно вычислить, не зная соответствующий секретный ключ. Здесь прослеживается аналогия с электронной подписью. Но отпечатки пальцев не являются подписями, поскольку могут быть «сняты», по крайней мере в принципе, без ведома клиента. Например, клиент может использовать свой секретный ключ для выполнения какого-либо протокола, предусмотренного процедурой купли-продажи. На основе транскрипции протокола продавец создает электронный отпечаток пальцев. При этом клиент может даже не подозревать, что полученная им копия файла содержит электронный отпечаток пальцев, и вообще не догадываться, что такие существуют в природе.

Проблемы, возникающие при попытке формализовать понятие электронного водяного знака, удобнее обсуждать на основе какого-либо определения, известного из литературы. Мы рассмотрим определение схемы электронных водяных знаков для защиты программного обеспечения из работы [6]. Под программой в этом определении понимается булева схема C (схема из функциональных элементов).

Электронный водяной знак обозначается через m и выбирается из некоторого множества водяных знаков, имеющих длину, подходящую для программы C. Программа, содержащая электронный водяной знак, называется помеченной.

Используется также следующее соглашение. Для функции $\alpha\colon \{0,1\}^* \to [0,1]$ формула $\forall\, x\ \alpha(x) = \nu(n)$ трактуется как сокращение следующего предложения: для любого полинома Q существует n_0: $\forall\, n \geqslant n_0\ \forall\, x \in \{0,1\}^n\ \alpha(x) < 1/Q(n)$.

Определение 16. Схемой электронных водяных знаков называется пара вероятностных алгоритмов $(\mathrm{Mark}, \mathrm{Extract})$ таких, что:

- (*функциональность*) Для всяких схемы C, ключа k и водяного знака m, строка $\mathrm{Mark}_k(C,m)$ описывает схему, которая вычисляет ту же функцию, что и C.
- (*эффективность помеченной программы*) Существует полином p такой, что для всякой схемы C, $|\,\mathrm{Mark}_k(C,m)| \leqslant p(|C| + |m| + |k|)$.
- (*извлекаемость*) Для всяких схемы C, ключа k, и водяного знака m, $\mathrm{Extract}_k(\mathrm{Mark}_k(C,m)) = m$.
- (*значимость*) Для всякой схемы C, $\Pr_k\{\mathrm{Extract}_k(C) \neq \lambda\} = \nu(|C|)$.
- (*хрупкость*) Для всякой полиномиальной вероятностной машины Тьюринга A существует полиномиальная вероятностная машина Тьюринга S такая, что для любых схемы C и водяного знака m

$$\left|\Pr_k\{A(\mathrm{Mark}_k(C,m)) = C' : C' \approx C \ \ \&\ \ \mathrm{Extract}_k(C') \neq m\}\right.$$
$$\left. -\ \Pr\left\{S^C(1^{|C|}) = C' : C' \approx C\right\}\right| = \nu(|C|).$$

Ключ k выбран наудачу из множества $\{0,1\}^{\max(|C|,|m|)}$, а $C' \approx C$ означает, что схемы C' и C функционально эквивалентны.

Напомним, что λ обозначает пустую строку.

Схема электронных водяных знаков называется эффективной, если функции Mark и Extract вычислимы за полиномиальное время.

В определении 16 свойство значимости отражает интуитивное требование, что большинство программ должны быть непомеченными (посредством водяных знаков). В самом деле, для любой программы C лишь ничтожная доля ключей k позволяет извлечь из программы что-либо отличное от пустой строки.

Из всех свойств, сформулированных в определении 16, основное внимание в литературе уделяется свойству хрупкости. В различных работах это свойство формулируется по-разному и даже неодинаково называется, но всегда формализует интуитивное требование стойкости: активный противник не может уничтожить водяной знак, не «разрушив» при этом весь файл. В том случае, когда электронные водяные знаки используются для защиты программного обеспечения, у противника всегда есть возможность проводить с программой эксперименты, получая пары (вход, выход). Если этой информации достаточно для того, чтобы построить программу, эквивалентную исходной, то водяные знаки как средство защиты не имеют смысла. Именно эта возможность учитывается в формулировке свойства хрупкости определения 16, где вероятность удаления противником водяного знака из помеченной программы $\mathrm{Mark}_k(C,M)$ сравнивается с вероятностью построения схемы, эквивалентной исходной схеме C, при доступе к последней как к оракулу.

И все же важнейшим свойством схемы электронных водяных знаков является извлекаемость. К сожалению, большинство авторов не уделяют этому свойству должного внимания. А ведь электронные водяные знаки представляют интерес только в том случае, если позволяют разрешать споры об интеллектуальной собственности. Предположим, что Боб торгует файлом F, а Алиса утверждает, что этот файл является ее интеллектуальной собственностью. Для разрешения спора она обращается к арбитру, предъявляя ему «полиномиальный вероятностный алгоритм» Extract и ключ k. Популярная теоретическая конструкция водяных знаков основывается на кодах, исправляющих ошибки, так что ключ k вполне может быть матрицей размера, скажем $10^5 \times 10^6$. Алгоритм Extract, проработав на

входных данных F, k несколько часов (полиномиальное время!), выдаст водяной знак Алисы, содержащий ее Copyright. Насколько такое доказательство будет убедительным для арбитра? И что должен делать арбитр, если Боб предоставит свой ключ k', с помощью которого другой (или даже тот же самый) алгоритм Extract извлечет из файла F его (Боба) водяной знак?

Процедура извлечения водяного знака должна быть простой и понятной. Например, если файл F содержит графическую информацию, то такая процедура, в принципе, могла бы выглядеть следующим образом. Алиса заявляет арбитру: «Замените все зеленые пиксели на черные, синие — на белые, и т. д., и посмотрите в левый нижний угол картинки. Вы увидите мой автограф.» Вероятно, такого рода доказательства могут быть убедительными для арбитра.

Ввиду того, что, на наш взгляд, не решена главная проблема на пути к адекватной формализации понятия электронного водяного знака (по крайней мере, авторам такое решение не известно), в данном разделе мы приводим единственный результат все из той же работы [6]. Следует особо подчеркнуть, что поскольку этот результат отрицательный, все приведенные выше критические замечания не применимы к определению 16. Чем шире класс алгоритмов Extract, о которых говорится в свойстве извлекаемости, тем отрицательный результат сильнее.

Теорема 7 ([6]). *Если существуют односторонние функции, то схемы электронных водяных знаков (в смысле определения 16) не существуют.*

Следствие 1 ([6]). *Эффективные схемы электронных водяных знаков (в смысле определения 16) не существуют.*

Литература

[1] Варновский Н. П. *О теоретико-сложностном подходе к определению стойкости стеганографических систем.* Сб. трудов 4-ой международной конференции «Дискретные модели в теории управляющих систем», 19–25 июля 2000 г. М.: «МАКС Пресс». 15–16.

[2] Anderson R. *Stretching the limits of steganography.* Proc. 1st Intern. Workshop on Inform. Hiding, 1996, LNCS, **v. 1174**, 39–48.

[3] Anderson R., Petitcolas F. *On the limits of steganography.* IEEE J. on Selected Areas in Communications, 1998, **v. 16**, № 4, 474–481.

[4] Cachin C. *An information-theoretic model for steganography.* Proc. 2nd Intern. Workshop on Inform. Hiding, 1998, LNCS, **v. 1525**, 306–318.

[5] Craver S. *On public-key steganography in the presence of an active warden.* Proc. 2nd Intern. Workshop on Inform. Hiding, 1996, LNCS, **v. 1525**, 355–368.

[6] Barak B., Goldreich O., Impagliazzo R., Rudich S., Sahai A., Vadhan S., Ke Yang. *On the (im)possibility of obfuscating programs.* J. Kilian (ed.). Advances in Cryptology — Crypto'01. LNCS, **v. 2139**, Springer-Verlag, 2001, 1–18. См. электронную версию данной работы: `http://www.math.ias.edu/~boaz/Papers/obfuscate.ps`.

[7] Diffie W., Hellman M. *New directions in cryptography.* IEEE Transactions om Information Theory, IT-22(6), 1976, 644–654.

[8] Dedić N., Itkis G., Reyzin L., Russell S. *Upper and lower bounds on black-box steganography.* Cryptology ePrint Archive (`http//eprint.iacr.org`), Report 2004/246.

[9] Dodis Y., Ong Sh. J., Prabhakaran M., Sahai A. *On the (im)possibility of cryptography with imperfect randomness.* Proc. 45th Symp. Found. of Computer Science, 2004, 196–205.

[10] Impagliazzo R., Levin L., Luby M. *Pseudo-random generation from one-way functions.* Proc. 21st Symp. on Theory of Computing, 1989, 12–24.

[11] Johnson N. F., Jajodia S. *Steganalysis of images created using current steganography software.* Proc. 2nd Intern. Workshop on Inform. Hiding, 1998, LNCS, **v. 1525**, 273–289.

[12] Luby M. *Pseudorandomness and cryptographic applications*. Princeton, NJ, Princeton University Press, 1996.

[13] Pfitzmann B. *Information hiding terminology*. Proc. 1st Intern. Workshop on Inform. Hiding, 1996, LNCS, **v. 1174**, 347–350.

[14] Simmons G. J. *The prisoners' problem and the subliminal channel*. Crypto'83, 1984, 51–67.

[15] Zöllner J., Federrath H., Klimant H., Pfitzmann A., Piotraschke R., Westfeld A., Wicke G., Wolf G. *Modeling the security of steganographic systems*. Proc. 2nd Intern. Workshop on Inform. Hiding, 1998, LNCS, **v. 1525**, 344–354.