Компьютерная математика

УДК 519.22; 004.415.24

Н.В. КОШКИНА

АНАЛИЗ БЕЗОПАСНОСТИ СИСТЕМ ЦИФРОВЫХ ВОДЯНЫХ ЗНАКОВ

Исследуется одна из базовых характеристик систем ЦВЗ – безопасность. Показаны ее отличия от более изученной характеристики – стойкости. Рассмотрены фундаментальные аспекты безопасности систем ЦВЗ, которые служат основой для дальнейшей разработки и исследования практических схем. Представлены две математические модели для анализа безопасности – информационно-теоретическая и вычислительная.

Введение. Проблематика создания цифровых водяных знаков (ЦВЗ) возникла в 90-х годах XX века как одно из направлений развития технологий информационной безопасности. Она обычно рассматривается в рамках науки о скрытой передаче информации – стеганографии [1 – 4]. Цифровой водяной знак представляет собой некоторые малообъемные специальные данные, внедряемые в сигнал-контейнер без нарушения его функциональности. Сигналом-контейнером при этом, как правило, выступают изображения, аудио- или видеосигналы. Технологии ЦВЗ находят свое применение при решении задач помехоустойчивой аутентификации данных (контроль целостности снимков камер наблюдения, записей телефонных разговоров, важной информации, хранящейся на бумажных носителях и т.п.), их источника (проверка того, что источник информации совпадает с заявленным) или владельца (защита авторских прав и прав собственности). Также они могут использоваться при контроле распространения сигналов, идентификации цифровых копий, скрытой аннотации данных (например, медицинских снимков, карт, музыки), контроле телевизионного и радиовещания, устройств копирования и др.

Разнообразие сфер применения технологий ЦВЗ повлекло за собой наличие разных толкований используемых понятий и оценок. Исследователями предлагается разная базовая терминология, которая как в любой молодой, «не устоявшейся» науке с течением времени претерпевает некоторые изменения и уточнения. Требования, предъявляемые к системе ЦВЗ, также могут достаточно сильно варьироваться в зависимости от ее применения.

Понятия стойкости и безопасности систем ЦВЗ

В начале пути становления и развития технологий ЦВЗ основное внимание исследователей уделялось усовершенствованию *стойкости* (robustness) цифрового водяного знака к естественным искажениям в стеганоканале: шуму, сжатию с потерями, фильтрации, геометрическим преобразованиям, ЦАП-АЦП и т.д. Естественные искажения происходят при обработке сигнала-контейнера и не связаны с эксплуатацией системы ЦВЗ, для которой они по сути являются неумышленными атаками. Обеспечение стойкости к естественным искажениям наряду с *неощутимостью* (imperceptibility) и *вместимостью* (capacity) образуемого стеганоканала – первая из проблем, с которой сталкиваются разработчики системы ЦВЗ.

Большинство применений технологий ЦВЗ подразумевает наличие легальных пользователей системы и нарушителей, которые имеют доступ к сигналу-контейнеру и могут предпринять умышленные атаки на систему – пассивную, активную или злоумышленную. Пассивная атака предполагает определение факта использования системы ЦВЗ и чтение внедренных данных. Активный нарушитель способен внести в маркированный контейнер модификации, не нарушающие его функциональность, но удаляющие ЦВЗ. Злоумышленник – внедрить фальшивый ЦВЗ в некоторый сигнал-контейнер.

Эксплуатация системы ЦВЗ в большинстве случаев не скрывается, а сама атака с целью определения факта наличия стеганоканала более характерна для систем скрытой передачи информации. При этом возможность чтения внедренных данных нелегальным пользователем, как правило, нежелательна, так как предоставляет ему знания для дальнейших активных или злоумышленных атак.

Удаление нарушителем ЦВЗ без получения им каких либо знаний о секретных параметрах системы принято называть слепым [5]. Такой подход реализован, например, в программе Stirmark [6]. Но очевидно, что при решении задач аутентификации или контроля копирования больший вред системе может быть нанесен не искажениями в канале или слепым удалением ЦВЗ, а нарушителем, добившимся возможности выполнять функции легального пользователя – обладателя секретного ключа. Единожды получив достоверную оценку секретного ключа, нарушитель может использовать ее для реализации атак многократно. Это послужило толчком для активизации интереса исследователей к еще одной базовой характеристике систем ЦВЗ – *безопасности* (security).

Следует отметить, что на сегодня в научном сообществе не определено четкое различие между такими характеристиками систем ЦВЗ как стойкость и безопасность. Достаточно часто в публикациях понятия стойкость и безопасность употребляются взаимозаменяемо или же, как например в [1], при рассмотрении всех возможных атак фигурирует термин стойкость. На различении рассматриваемых характеристик систем ЦВЗ заострено внимание в работах [5, 7, 8].

Исследование проблематики стойкости систем ЦВЗ подразумевает оценку стойкости цифрового водяного знака к искажениям, умышленно или неумышленно привнесенным в маркированный сигнал-контейнер. Стойкость ЦВЗ

оценивается количеством ошибок, возникающих при его извлечении легальным пользователем из искаженного контейнера. Необходимый уровень стойкости определяется применением системы. Системы ЦВЗ, применяемые для защиты авторского права, аутентификации источника данных, контроля распространения или копирования, должны обеспечивать стойкость ЦВЗ ко всем возможным искажениям маркированного сигнала-контейнера. При этом системы ЦВЗ, применяемые для аутентификации данных, должны или не позволять вообще никаких модификаций маркированного контейнера, или обеспечивать стойкость ЦВЗ только к допустимым искажениям: например, позволять сжатие с потерями, но не позволять вставку / удаление фрагментов. То есть, при рассмотрении проблематики стойкости систем ЦВЗ выделяют системы со стойкими, хрупкими и полухрупкими водяными знаками [1].

Безопасность непосредственно связана со сложностью получения оценки секретных параметров системы, т. е. синонимом безопасности системы ЦВЗ можно считать стойкость к получению нарушителем секретов системы, что с учетом принципа Керкгоффа выливается в стойкость к восстановлению секретных ключей. Ключи системы ЦВЗ параметризируют функции внедрения и извлечения. Они могут определять область внедрения (временая/пространственная или частотная), базис частотного разложения, правила разбиения контейнера на сегменты, силу внедрения, индексы маркируемых коэффициентов, точки квантования, используемую кодовую книгу, расширяющий вектор и т. п. В дальнейшем будем предполагать систему ЦВЗ симметричной, т. е. при извлечении водяного знака используется тот же секретный ключ, что и при его внедрении.

Необходимый уровень безопасности также определяется применением системы. Например, системы ЦВЗ, применяемые для скрытой аннотации данных с целью удобства организации их хранения и поиска, не предполагают наличие нарушителя, их задача сосредотачивается на обеспечении стойкости ЦВЗ к операциям обработки контейнера. При этом системы, применяемые для аутентификации, контроля распространения или мониторинга должны ориентироваться на существование потенциальных нарушителей, следовательно их разработка предполагает максимизацию достижимого уровня безопасности.

Атаки, угрожающие безопасности, всегда являются умышленными и не слепыми, но не все умышленные или не слепые атаки нацелены на нарушение безопасности. Информация, полученная посредством атак, угрожающих безопасности, может использоваться как первый шаг для выполнения атак, угрожающих стойкости.

Безопасность должна оцениваться независимо от стойкости. Здесь показательной является аналогия с криптографией: цель атакующего в криптографии – расшифровка секретного сообщения; безопасность системы оценивается в предположении безошибочности канала коммуникации, поскольку иначе сообщение будет разрушено и для нарушителя, и для легального пользователя. Для системы ЦВЗ это будет означать, что безопасность оценивается в предположении отсутствия атак, угрожающих стойкости.

Отметим, что в целом научных публикаций, рассматривающих вопрос стойкости, гораздо больше, чем работ, посвященных безопасности систем ЦВЗ. Первая попытка разделить данные понятия выполнена в [7] и в дальнейшем развита в работах [8, 9] и др. Исторический очерк формирования определения безопасности систем ЦВЗ можно найти в [5]. В отечественных публикациях эта тематика остается практически незатронутой. Как следствие, относительно многих предлагаемых методов сложно сказать достигают ли они поставленных целей, даже если проблема стойкости полностью решена, поскольку при их разработке не было уделено достаточно внимания проблеме безопасности. Рассмотрим теоретические основы безопасности систем ЦВЗ.

Информационно-теоретический подход к оценке безопасности систем ЦВЗ

Начиная с [10] для анализа безопасности стали использоваться информационно-теоретические модели, которые по аналогии с криптографией берут за основу меру Шеннона [11] и адаптируют ее под системы ЦВЗ. При таком подходе пустые контейнеры, ЦВЗ, ключи и маркированные контейнеры рассматриваются как случайные переменные. Обозначим их X, M, K и Y соответственно.

Внедрение ЦВЗ в сигнал-контейнер можно представить аддитивной моделью (рис. 1).

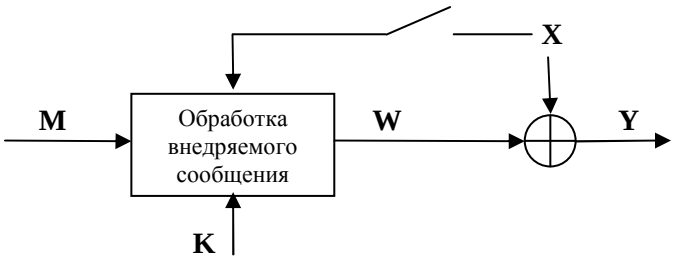

РИС. 1. Обобщенная модель функции внедрения

Пусть значением переменной X является некоторый сигнал-контейнер x, а значением исходного ЦВЗ M – сообщение m. Тогда m с помощью ключа k (реализация K) и, возможно, исходного контейнера x преобразовывается в ЦВЗ w. Сумма w и x дает итоговый маркированный контейнер y (реализация Y).

На практике легальный пользователь системы ЦВЗ обычно имеет свой собственный секретный ключ и использует его для маркировки цифровых данных многократно. Следовательно все маркированные одним пользователем контейнеры (или, по крайней мере, большое количество) будут содержать информацию об одном и том же секретном ключе. Информация о секретном ключе проявляется при наблюдениях и таким образом может стать доступной нарушителю. Как правило, для получения достоверной оценки ключа нужно иметь много маркированных им сигналов. Но как только такая оценка будет получена, она может быть использована для атак на большое количество контейнеров без каких либо дополнительных усилий.

Атаки, угрожающие безопасности, можно классифицировать, положив в основу содержимое доступных нарушителю наблюдений $O^{N_o} = O_1, O_2 \ldots O_{N_o}$. Например, нарушитель может знать имя автора в сценариях защиты авторского права или статус фильма в сценариях защиты от копирования; иметь DVD с маркированным фильмом и более раннюю версию фильма, которая не была защищена; использовать доступ к устройству внедрения или обнаружения для оценки чужого ключа аутентификации и др. Таким образом, независимо от практического применения системы ЦВЗ выделяют:

- атаку на основе маркированных контейнеров: у атакующего есть доступ только к маркированным контейнерам y^{N_O};
- атаку на основе известного ЦВЗ: у атакующего есть доступ к парам маркированный сигнал и соответствующий ему ЦВЗ $(y,m)^{N_O}$;
- атаку на основе известного оригинала: нарушитель проектирует свою атаку, базируясь на знании пар оригинальных и соответствующих им маркированных контейнеров $(x,y)^{N_O}$.

Отметим, что эта классификация не охватывает всех возможных применений технологий ЦВЗ и при необходимости может быть расширена добавлением других важных сценариев.

Пусть задан метод внедрения/извлечения ЦВЗ и псевдослучайно определен ключ. Нарушитель знает метод, но не знает ключ. Как случайная переменная секретный ключ обладает некоторой энтропией. Обозначим априорную энтропию ключа, т. е. энтропию до начала каких либо действий как $H(K)$. В дальнейшем продуцируется N_o маркированных контейнеров, с каждым из которых нарушителю просачивается немного информации. Доступность ЦВЗ или пустых контейнеров также добавляет нарушителю знаний. Полученную в итоге апостериорную энтропию ключа обозначим $H(K \mid O^{N_O})$. Тогда утечку информации, измеряемую взаимной информацией между выполненными наблюдениями и секретным ключом, можно выразить как разность между априорной и апостериорной энтропией ключа:

$$I(K; O^{N_O}) = H(K) - H(K \mid O^{N_O}). \qquad (1)$$

Для оценки безопасности системы ЦВЗ нужно знать как минимум две из трех величин в формуле (1). Чем больше утечка информации $I(K; O^{N_O})$, тем меньше неопределенность для нарушителя $H(K \mid O^{N_O})$, которая с ростом количества наблюдений монотонно спадает от $H(K)$ до 0. Когда $H(K \mid O^{N_O})$ становится нулевой, это означает, что нарушитель обладает достаточным для определения ключа количеством наблюдений.

Методика оценки безопасности систем ЦВЗ в информационно-теоретической модели базируется на следующих определениях.

Определение 1. Система ЦВЗ достигает абсолютной безопасности при $I(K;O^{N_o})=0$.

Это означает, что все усилия нарушителя, направленные на раскрытие секретного ключа будут бесполезны, даже если он обладает бесконечной вычислительной мощностью. Очевидно, что разработка систем ЦВЗ, для которых выполняется это определение, может быть чрезвычайно трудной задачей или приводить к непрактичным системам (например, из-за вычислительной сложности или длины ключа).

Так же в информационно-теоретической модели вводится понятие $N-\varepsilon$ безопасности.

Определение 2. Система ЦВЗ $N-\varepsilon$ безопасна, если $I(K;O^{N})\le\varepsilon$ для некоторой константы $\varepsilon>0$.

Отметим, что утечка информации может быть нулевой или маленькой вследствие нулевой или маленькой априорной энтропии секретного ключа и это необходимо учитывать при анализе безопасности системы. Тут показательным случаем есть граничный, т. е. рассмотрение детерминированного ключа. При детерминированном ключе утечка информации является нулевой, однако в силу отсутствия секретной параметризации система испытывает существенный недостаток безопасности.

Такой анализ дает начало понятию уровня безопасности, определяемому как более удобная мера.

Определение 3. Для систем с $I(K;O^{N_o})\neq 0$ под γ-уровнем безопасности понимают число наблюдений N_γ, необходимых для выполнения условия $\mathrm{H}(K\,|\,O^{N_\gamma})\le\gamma.$

Определение 4. Количество наблюдений N_υ, приводящих к детерминированному ключу, называют расстоянием уникальности.

В отличие от безопасности, стойкость выражается как $I(M;Y'\,|\,K)$, где Y' – множество искаженных контейнеров. Тогда проектирование систем ЦВЗ будет сопряжено с минимизацией $I(K;O^{N_o})$ – достижимого количества информации о секретном ключе для нарушителя и одновременной максимизацией $I(M;Y'\,|\,K)$ – достижимого количества информации для легального пользователя.

Отметим, что кроме представленного подхода на сегодня в литературе описана информационно-теоретическая модель, в которой рассматривается адаптация подхода Шеннона на случай непрерывных случайных переменных [5], что предполагает замену энтропии дифференциальной энтропией. Так же существует подход, в котором утечка информации измеряется с помощью информационной матрицы Фишера [12].

Вычислительный подход к оценке безопасности систем ЦВЗ

Информационно-теоретические модели показывают фундаментальные, безоговорочные слабости безопасности. Используя их для аппроксимации канала распространения, удалось аналитически определить уровень безопасности таких методов создания ЦВЗ, как расширение спектра и модуляция индекса квантования [13–14]. Однако они обладают несколькими очевидными недостатками – не учитывают вычислительную мощность атакующего и могут оказаться сложно реализуемыми на практике. Поэтому, как и в криптографии, параллельно с информационно-теоретическим развивается вычислительный (теоретико-сложносный) анализ безопасности [15].

Атакующий моделируется вероятностной машиной Тьюринга с полиномиальным ограничением на время вычислений. Во время своей работы он может принимать случайные решения, рассматриваемые как подбрасывания монеты. Цель атакующего – «получение знаний о секретном ключе» – моделируется способностью различить с каким из двух ключей в данный маркированный контейнер был внедрен ЦВЗ. Формально безопасность системы ЦВЗ определяется через игру между атакующим и арбитром, которому доверяют. Игра состоит из двух частей, первая представляет собой операции атакующего по сбору информации, вторая – проведение испытания:

1) арбитр генерирует два ключа $k_1, k_2 \in K$ длины n и предоставляет доступ атакующему к двум оракулам Θ_{k_1} и Θ_{k_2}, осуществляющим внедрение водяного знака с ключом k_1 и k_2 соответственно. Атакующий (адаптивно) продуцирует тестовые сигналы-контейнеры $x_1, x_2 \ldots \in X$ и, используя оракулы по собственному выбору, получает их маркированные версии $y_1, y_2 \ldots \in Y$. Атакующий может свободно выполнять вероятностные полиномиальные по времени вычисления над маркированными контейнерами;

2) когда атакующий закончит процесс сборки информации, арбитр подбрасывает монету $b \in \{0,1\}$, продуцирует объект y_d с ключом k_b и передает y_d атакующему. Последний, не имея больше доступа к оракулам, должен определить с каким ключом был получен y_d – k_1 или k_2. То есть, он должен определить бит b, случайно выбранный арбитром. На этом шаге атакующий также может выполнять вероятностные полиномиальные операции.

Преимущество атакующего в выигрыше игры может использоваться для оценки безопасности системы ЦВЗ. Это преимущество определяется как вероятность правильного предположения атакующим бита b минус 0.5 и измеряет систематический шанс различить, с каким из двух ключей был внедрен ЦВЗ в данный маркированный контейнер. Отметим, что атакующий всегда может сделать случайный выбор и выиграть игру с вероятностью 0.5. Абсолютно безопасная система ЦВЗ должна приближать преимущество атакующего к 0. В идеале преимущество должно слабо зависеть от длины ключа.

В вышеприведенном сценарии игры атакующий не может выбрать контейнер y_d, получаемый им на этапе испытания. Если по аналогии атак с выбранным открытым текстом в криптографии позволить атаки с выбранным исходным контейнером, вторая часть сценария игры модифицируется следующим образом:

2') когда атакующий закончит процесс сборки информации, он вычисляет тестовый контейнер x, отличающийся от всех предыдущих запросов оракулу $x_1, x_2 \ldots$, и передает x арбитру, который подбрасывает монету $b \in \{0,1\}$ и внедряет ЦВЗ с ключом k_b в сигнал x, результатом чего будет y_d. Арбитр передает y_d атакующему. Последний, не имея больше доступа к оракулам, должен определить с каким ключом был получен y_d – k_1 или k_2.

В варианте описанной игры 2' получим более сильное понятие безопасности, чем в 2.

В некоторых практических сценариях, как например, защита от копирования CD и DVD дисков, нарушитель может воспользоваться доступом к детектору ЦВЗ в виде «черного ящика». Для получения знаний о секретном ключе во время такой атаки он делает запросы детектору ЦВЗ с целенаправленно измененными тестовыми сигналами.

В этом случае знание о секретном ключе моделируется полиномиально вычислимым предикатом $P: \{0,1\}^* \to \{0,1\}$, который отображает двоичный секретный ключ в бит b, т. е. способностью определить, обладает секретный ключ некоторым бинарным свойством или нет. Безопасность систем ЦВЗ с общедоступным детектором оценивается через игру между атакующим и арбитром, где задача атакующего вычислить свойство секретного ключа $P(K)$ с помощью оракула детектирования содержащего данный ключ:

1) арбитр генерирует случайный ключ $k \in K$ длины n и ЦВЗ $m \in M$. Далее он предоставляет доступ атакующему к оракулу детектирования Θ_k, который возвращает «true» тогда и только тогда, когда m обнаруживается во входном сигнале-контейнере;

2) атакующий неоднократно генерирует тестовые сигналы $y_1, y_2 \ldots$ и может использовать оракула детектирования для проверки наличия в них ключа. Результатом его полиномиально ограниченных вычислений должен быть бит b. Атакующий выигрывает игру, если $b = P(K)$, т. е. если он правильно предполагает свойство использованного ключа.

Заключение. Четкое определение понятия безопасность систем ЦВЗ важно для разработки эффективных практических методов и алгоритмов. Формальные модели безопасности позволяют оценивать и сравнивать безопасность разных методов, а также строить доказуемо безопасные системы.

В работе рассмотрена проблематика безопасности систем ЦВЗ, в которых один и тот же ключ многократно используется для маркировки сигналов-контейнеров, что актуально для широкого круга приложений. Рассмотрены базовые математические модели для анализа безопасности. Системы ЦВЗ с низкой согласно рассмотренным подходам безопасностью, могут быть безопасными в приложениях, где при каждом внедрении ЦВЗ используется новый секретный ключ.

В качестве направления будущих исследований отметим дальнейшее развитие рассмотренных подходов, анализ безопасности различных методов внедрения ЦВЗ в описанных математических моделях. Так же перспективной с точки зрения защиты мультимедийных данных представляется интеграция криптографических технологий и ЦВЗ.

Н.В. Кошкіна

АНАЛІЗ БЕЗПЕКИ СИСТЕМ ЦИФРОВИХ ВОДЯНИХ ЗНАКІВ

Досліджується одна з базових характеристик систем ЦВЗ – безпека. Показано її відмінності від більш вивченої характеристики – стійкості. Розглянуто фундаментальні аспекти безпеки систем ЦВЗ, які є основою для подальшої розробки та дослідження практичних схем. Представлені дві математичні моделі для аналізу безпеки – інформаційно-теоретична та обчислювальна.

N.V. Koshkina

SECURITY ANALYSIS OF WATERMARKING SYSTEMS

We investigate one of the basic feachures of watermarking systems, namely, security. A distinction of it from a robustness (the most studied characteristic) is shown. Fundamental aspects of security watermarking systems, which form a basis for further working out and investigation of practical schemes are considered. Two mathematical models for security analysis (information-theoretical and computational) are presented in the paper.

1. *Грибунин В.Г., Оков И.Н., Туринцев И.В.* Цифровая стеганография. – М.: СОЛОН-Пресс, 2002. – 261 с.
2. *Аграновский А.В., Девянин П.Н., Хади Р.А., Черемушкин А.В.* Основы компьютерной стеганографии. – М.: Радио и связь, 2003. – 152 с.
3. *Хорошко В.А., Шелест М.Е.* Введение в компьютерную стеганографию. – Киев: Національний авіаційний університет, 2002. – 152 с.
4. *Конахович Г.Ф., Пузыренко А.Ю.* Компьютерная стеганография. Теория и практика.– Киев: МК-Пресс, 2006. – 288 с.
5. *Perez-Freire L., Comesana P., Troncoso-Pastoriza J.R., Perez-Gonzalez F.* Watermarking security: a survey // Transactions on Data Hiding and Multimedia Security, 2006. – **4300**. – P. 41–72.
6. *Petitcolas F.A.P.* Stirmarkbenchmark 4.0, режим доступа: http://www.petitcolas.net/fabien/watermarking/stirmark/

7. *Kalker T.* Considerations on watermarking security // IEEE International Workshopon Multimedia Signal Processing. – Cannes(France), 2001. – P. 201–206.
8. *Cayre F., Fontaine C., Furon T.* Watermarking security: Theory and practice // IEEE Trans. Signal Processing, 2005. – **53**. – P. 3976–3987.
9. *Furon T.* A survey of watermarking security // Proc. of Int. Work. on Digital Watermarking. – Siena (Italy), 2005. – **3710**. – P. 201—215.
10. *Mitthelholzer T.* An information-theoretic approach to steganography and watermarking // 3rd Int. Workshop on Information Hiding. – Dresden (Germany), 1999. – **1768**. – P. 1–17.
11. *Shannon C.E.* Communication theory of secrecy systems // Bell system technical journal, 1949. – **28**. – P. 656–715.
12. *Fisher R.A.* On the mathematical foundations of theoretical statistics // Philosophical Transactions of the Royal Society, 1922. – **222**. – P. 309–368.
13. *Comesana P., Perez-Freire L., Perez-Gonzalez F.* Fundamentals of data hiding security and their application to spread-spectrum analysis // Information Hiding: 7th international workshop. – Barcelona (Spain), 2005. – **3727**. – P. 146–160.
14. *Perez-Freire L., Perez-Gonzalez F., Comesana P.* Secret dither estimation in lattice-quantization data hiding: a set-membership approach // Security, Steganography, and Watermarking of Multimedia Contents VIII. – San Jose (California, USA), 2006. – **6072**. – P. 1–12.
15. *Katzenbeisser S.* Computational security models for digital watermarks // In 6th International Workshop on Image Analysis for Multimedia Interactive Services. – Montreux (Switzerland), 2005.

Получено 25.11.2010

Об авторе:

Кошкина Наталия Васильевна,
кандидат физико-математических наук, старший научный сотрудник
Института кибернетики имени В.М. Глушкова НАН Украины.
K_n_v@ukr.net