СТЕГОСИСТЕМЫ ЦИФРОВЫХ ВОДЯНЫХ ЗНАКОВ

© НАЗАРОВА Валентина Ивановна
преподаватель гимназии № 1599, г. Москва.
☎ (499) 172-06-05

Статья посвящена одному из современных направлений защиты информации от несанкционированного доступа – стеганографии и её разделу – встраиванию цифровых водяных знаков в цифровые фото-аудио-видео данные. Рассматривается принципиальная математическая модель стегосистемы цифровых водяных знаков, приведена классификация подобных стегосистем.

Ключевые слова: *защита информации, стеганография, цифровые водяные знаки, математическая модель.*

Задача *защиты информации от несанкционированного доступа* была актуальной на протяжении всей истории существования человечества. Ещё в древнем мире выделилось два основных направления в этой области: *криптография* и *стеганография*. Целью криптографии является скрытие содержимого сообщений за счёт их шифрования. Стеганографиия предполагает скрытие самого факта существования сообщения. Эти направления существуют и динамично развиваются и в настоящее время.

Совершенствование средств вычислительной техники в последние десятилетия дало новый толчок для развития стеганографии. Секретные сообщения встраивают в цифровые данные, как правило, имеющие аналоговую природу, – речь, аудиозаписи, изображения, видео. Известны также предложения по встраиванию секретной информации в текстовые файлы и даже в исполняемые файлы программ.

В истории компьютерной стеганографии были два последовательных этапа. Первый из них не связан с цифровой обработкой сигналов. В этом случае секретное сообщение может быть встроено в заголовки файлов, пакетов данных. Такой способ не получил распространения в связи с относительной лёгкостью вскрытия и/или уничтожения скрытой информации. Большинство текущих исследований в области стеганографии так или иначе связаны со вторым этапом – *применением цифровой обработки сигналов*. Накопление научных знаний в этом векторе развития способствовало появлению *цифровой стеноргафии* как науки. Она включает в себя следующие четыре *основных направления*:

1) встраивание информации с целью её скрытой передачи;

2) встраивание цифровых водяных знаков[1] (watermarking);

3) встраивание идентификационных номеров (fingerprinting);

4) встраивание заголовков (captioning).

Остановимся подробнее на втором направлении. *ЦВЗ применяются, в основном, для защиты от копирования и несанкционированного использования цифровых данных аналоговой природы*. Примерами могут послужить фотографии, аудио и видеозаписи и т. д. В связи с бурным развитием цифровых мультимедийных технологий остро встал *вопрос защиты авторских прав и интеллектуальной собственности на произведения, представленные в цифровом виде*. Преимущества представления и передачи цифровых фото-аудио-видеоданных при помощи средств компьютерной техники могут

[1] Далее по тексту – ЦВЗ.

оказаться перечёркнутыми лёгкостью, с которой возможно воровство или несанкционированная модификация указанных данных. Поэтому разрабатываются различные меры защиты информации, носящие как организационный, так и технический характер. Один из наиболее эффективных *приёмов защиты мультимедийной информации* заключается во встраивании в защищаемый объект невидимых меток – ЦВЗ. Разработки в этой области ведут крупнейшие фирмы во всем мире. Так как Методы ЦВЗ начали разрабатываться относительно недавно (первой работой на эту тему был доклад японских математиков K. Matsui, K. Tanaka и Y. Nakamura на Симпозиуме по криптографии и информационной безопасности в 1989 г.), то здесь имеется много неясностей и проблем, требующих своего разрешения.

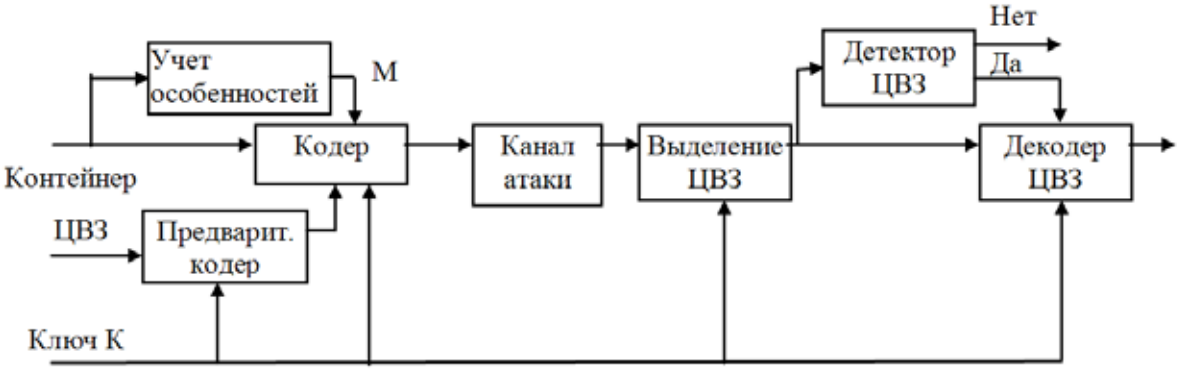

Рис. 1. Структурная схема типичной стегосистемы ЦВЗ

Задачу встраивания и выделения ЦВЗ из другой информации выполняет стегосистема ЦВЗ, состоящая из следующих основных элементов (рис. 1):

предварительный кодер – устройство, предназначенное для преобразования скрываемого сообщения к виду, удобному для встраивания в сигнал-контейнер[1];

кодер – устройство, предназначенное для осуществления вложения скрытого сообщения в другие данные с учётом их модели;

выделение ЦВЗ – устройство выделения встроенного сообщения;

детектор ЦВЗ – устройство, предназначенное для определения наличия встроенного сообщения;

декодер ЦВЗ – устройство, восстанавливающее скрытое сообщение.

В стегосистеме ЦВЗ происходит объединение двух типов информации – открытого сообщения и скрытого сообщения – в едином стегосообщении так, чтобы они могли быть различимы двумя разными детекторами. В качестве одного из детекторов выступает система выделения ЦВЗ (на схеме и далее по тексту – выделение ЦВЗ). В качестве другого детектора как правило выступает человек (на схеме и далее по тексту – детектор ЦВЗ).

Рассмотрим *математическую модель стегосистемы ЦВЗ*.

Стегосистему можно представить как систему связи [1]. Сам *алгоритм встраивания ЦВЗ* состоит из трёх основных этапов:

1) генерации ЦВЗ,

2) встраивания ЦВЗ в кодере;

3) обнаружения ЦВЗ в детекторе.

Приведём формальное математическое описание каждого из этапов.

Генерация ЦВЗ.

Пусть W^*, K^*, I^*, B^* есть множества возможных ЦВЗ, ключей, контейнеров и скрываемых сообщений соответственно. Тогда генерация ЦВЗ может быть представлена в виде функции:

$$F: \ I^* \times K^* \times B^* \to W^*, \ W = F(I, K, B), \quad (1)$$

где W, K, I, B - представители соответствующих множеств.

Функция F может быть произвольной, но на практике требования робастности[2] ЦВЗ накладывают на неё ограничения. Так, в большинстве случаев должно выполняться требование:

$$F(I, K, B) \approx F(I + \varepsilon, K, B), \quad (2)$$

то есть незначительно изменённый контейнер не приводит к изменению ЦВЗ.

Функция F обычно является составной:

$$F = T \circ G,$$

[1] Контейнером называется информационная последовательность, в которой прячется сообщение.

[2] Под робастностью понимается устойчивость ЦВЗ к различного рода внешним воздействиям. Робастные ЦВЗ являются наиболее практичными, и большинство разработок в рассматриваемой области посвящено именно робастным ЦВЗ.

где $G: K^* \times B^* \to C^*$ и $T: C^* \times I^* \to W^*$, (3)

следовательно, ЦВЗ зависит от свойств контейнера.

Функция G может быть реализована при помощи криптографически безопасного генератора псевдослучайных последовательностей с K в качестве начального значения. Функция T преобразует кодовые слова C^*, в результате чего получается ЦВЗ W^*. На эту функцию можно не накладывать ограничения необратимости, так как соответствующий выбор G влечёт необратимость F. Однако функция T должна быть выбрана так, чтобы незаполненный контейнер I_o, заполненный контейнер I_W и незначительно модифицированный заполненный контейнер I'_W порождали бы один и тот же ЦВЗ:

$$T(C, I_0) = T(C, I_W) = T(C, I'_W) \quad (4)$$

то есть она должна быть устойчивой к незначительным искажениям контейнера.

Встраивание ЦВЗ. Этот процесс происходит в кодере и может быть описан как *суперпозиция двух сигналов:*

$$I_w(i,j) = I_0(i,j) \oplus L(i,j)\, W(i,j)\, p(i,j), \quad (5)$$

где $W(i,j)$ – ЦВЗ;

$I_o(i,j)$ – исходное сообщение;

$L(i,j)$ – маска встраивания ЦВЗ, служащая для уменьшения их заметности и учитывающая характеристики детектора ЦВЗ[1];

$p(i,j)$ – проектирующая функция, зависящая от ключа;

знаком $\oplus$ обозначен оператор суперпозиции.

Проектирующая функция осуществляет «распределение» ЦВЗ по области изображения. Её применение можно представить в виде разнесения информации по параллельным каналам. Кроме того, эта функция имеет определённую структуру и свойства, использующиеся для противодействия атакам.

Другое возможное описание процесса внедрения ЦВЗ может быть получено при *представлении стегосистемы как системы связи с передачей дополнительной информации* (рис. 2) [1].

В этой модели кодер и декодер имеют доступ, помимо ключа, к полной информации о канале (то есть о контейнере и возможных атаках). Степень доступа регулируется переключателями A и B. В зависимости от их положения выделяют *четыре класса стегосистем.*

I класс: дополнительная информация отсутствует (переключатели разомкнуты) – это

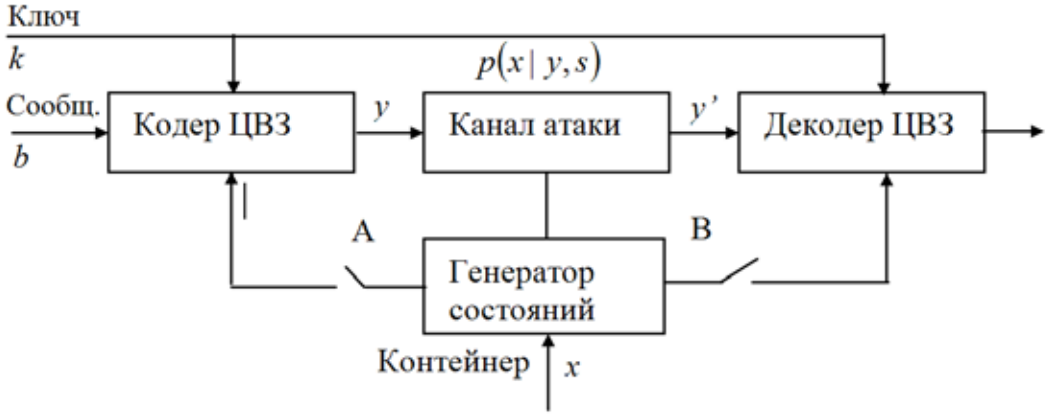

Рис. 2. Представление стегосистемы, как системы связи с передачей дополнительной информации

так называемые «классические» стегосистемы. Обнаружение ЦВЗ осуществляется путём вычисления коэффициента корреляции между принятым стегосообщением и вычисленным по ключу ЦВЗ. Если коэффициент превышает некоторый порог, выносится решение о присутствии ЦВЗ. На практике данные системы не являются эффективными[2].

II класс: информация о канале известна только кодеру (A замкнут, B разомкнут) [2]. Достоинством данной схемы является то, что она имеет в теории ту же пропускную способность, что и схема с наличием связи исходного контейнера с декодером. К недостаткам стегосистем этого класса можно отнести высокую сложность организации кодера (необходимость построения кодовой книги для каждого сообщения), а также отсутствие адаптации схемы к возможным атакам. Предложен ряд практических подходов,

[1] В большинстве случаев маска встраивания ЦВЗ должна ориентироваться на особенности зрительной системы человека

[2] Корреляционный при`мник оптимален лишь в случае аддитивной гауссовой помехи. При других атаках (например, геометрических искажениях) эти стегосистемы показывают неудовлетворительные результаты.

преодолевающих эти недостатки. В частности, для снижения сложности кодера предлагается использовать структурированные кодовые книги.

III класс: дополнительная информация известна только декодеру (*A* разомкнут, *B* замкнут). В таких схемах декодер строится с учётом возможных атак, в результате чего получаются робастные к различным атакам системы. Для достижения этой цели применяются разные методы, в частности использование опорного ЦВЗ[1]. Например, можно выполнить встраивание в амплитудные коэффициенты преобразования Фурье, которые инвариатны к аффинным преобразованиям. В этом случае опорный ЦВЗ «покажет», какое преобразование выполнил со стегосообщением атакующий.

IV класс: дополнительная информация известна и в кодере, и в декодере (оба ключа замкнуты). Все перспективные стегосистемы должны строиться по этому принципу [3]. Эффективность этой схемы достигается путём согласования кодера с сигналом-контейнером, а также адаптивным управлением декодером в условиях наблюдения канала атак.

Обнаружение ЦВЗ в детекторе.

В зависимости от типа стегодетектора он может выдавать двоичные либо *M*-ичные решения о наличии/отсутствии ЦВЗ. В первом случае детектор называется «жёстким», во втором – «мягким».

Рассмотрим вначале более простой случай *«жесткого» детектора*.

Обозначим операцию детектирования через *D*, тогда:

$$D : I_W {}^* \times K^* \to \{0,1\},$$

$$D(I_W, W) = D\left(I_W, F(I_W, K)\right) = \begin{Bmatrix} 1, & \text{если } W \text{ есть} \\ 0, & \text{если } W \text{ нет} \end{Bmatrix}. \quad (6)$$

В качестве детектора ЦВЗ обычно используют *корреляционный приёмник* (рис. 3).

Работу корреляционного детектора рассмотрим на примере, когда сообщением выступает растровое изображение. Пусть у половины пикселов изображения значение яркости увеличено на 1, а у остальных не изменилось либо уменьшилось на 1. Тогда $I_W=I_o+W$, где $F(I_o,K)=W$.

Коррелятор детектора ЦВЗ вычисляет величину $I_W \cdot W=(I_o+W)\cdot W=I_o\cdot W+W\cdot W$. Так как W может принимать значения ±1, то $I_o \cdot W$ будет весьма мало, а $W\cdot W$ – всегда положительно. Поэтому $I_W \cdot W$ будет близко к $W\cdot W$. Теперь можно записать вероятность неверного обнаружения стегосообщения как дополнительную (комплементарную) функцию ошибок от корня квадратного из отношения $W\cdot W$ (так называемая «энергии сигнала») к дисперсии значений пикселов яркости (так называемая «энергия шума»).

Для случая мягкого детектора и закрытой стегосистемы имеем две основные меры похожести: нормированный коэффициент взаимной корреляции $\delta = \frac{I_0 I_W}{\|I_0\|\|I_W\|}$ и расстояние по *Хэммингу* $\delta = N - \sum_{i=1}^{N} i_0 i_W$.

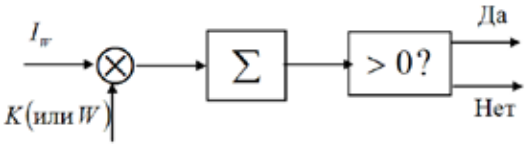

Рис. 3. Корреляционный детектор ЦВЗ

В таком детекторе возможно возникновение двух типов ошибок:

1) существует вероятность того, что детектор не обнаружит имеющийся ЦВЗ;

2) существует вероятность ложного нахождения ЦВЗ в пустом контейнере (вероятность ложного обнаружения).

Снижение одной вероятности приводит к увеличению другой. Надежность работы такого детектора характеризуется вероятностью ложного обнаружения. Система ЦВЗ должна быть построена таким образом, чтобы минимизировать вероятности возникновения обеих ошибок, так как каждая из них может привести к отказу от обслуживания.

Библиографический список

1. **Voloshynovskiy S., Pereira S., Iquise V., Pun T**. Attack Modelling: Towards a Second Generation Watermarking Benchmark // Preprint. University of Geneva. – 2001. – 58 p.

2. **Cox J., Miller M., McKellips A.** Watermarking as communications with side information // Proceedings of the IEEE. – 1999. Vol. 87. – № 7. – P. 1127–1141.

3. **Marvel L.** Image Steganography for hidden communication. PhD Thesis Univ. of Delaware, 1999. – 115 p.

[1] Опорный ЦВЗ представляет собой небольшое число бит, внедряемых в инвариантные к внешним преобразованиям коэффициенты сигнала.