МИНОБРНАУКИ РОССИИ
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«САРАТОВСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ИМЕНИ Н.Г. ЧЕРНЫШЕВСКОГО»

Кафедра теоретических основ компьютерной безопасности и криптографии

Исследование стойкости ЦВЗ, встроенных методом DEW, к атакам

АВТОРЕФЕРАТ
дипломной работы

студентки 6 курса 631 группы
специальности 10.05.01 Компьютерная безопасность
факультета компьютерных наук и информационных технологий
Лукьяновой Анны Алексеевны

Научный руководитель
доцент, к. п. н. _____ А. С. Гераськин
23.01.2021 г.

Заведующий кафедрой
д. ф.-м. н., доцент _____ М. Б. Абросимов
23.01.2021 г.

Саратов 2021

ВВЕДЕНИЕ

Методы стеганографии использовались человечеством еще задолго до появления компьютеров. Геродот в своем трактате «История», относящемся к 440 году до н.э., описывает следующий метод: сообщение записывалось на деревянную подложку восковой таблички еще до нанесения воска[1]. С появлением фотографии стали доступны новые способы стеганографии – например, в самые обычные снимки добавлялись микроточки. Подобные методы передачи секретных сообщений активно использовались в период Второй мировой войны[2]. Можно найти и массу иных примеров стеганографии в прошлом[3].

Вопрос обеспечения защиты передаваемой информации актуален и сегодня. В связи с быстрым развитием интернета и ростом пропускной способности каналов передачи, появилась возможность размещения на сетевых ресурсах информации в форматах, отличных от текстового, например, фото, видео и аудио. Данные возможности имеют свои преимущества и недостатки. В связи с такой доступностью, и простотой размещения материалов, появляется возможность их копирования и использования в своих целях. Поэтому, наряду с удобством публикации и продвижения произведений графической информации в интернете, появляется потребность в защите прав на эти изображения[4]. При этом проблема осложняется тем, что во многих странах существуют законы, запрещающие применение стойких криптографических алгоритмов. На применение стеганографии подобных ограничений пока не наложено. Именно это делает привлекательным ее использование как дополнительного способа

---

[1] Petitcolas, F.A.P., Anderson, G.J., Kuhn, M.G. Information Hiding – A Survey / Proceedings of IEEE: Special issue on protection of multimedia content. - 1999. - vol. 87, no. 7. - pp. 1062-1078

[2] Баранов, С.А., Голодков, Ю.Э., Демаков, В.И., Кургалеева, Е.Е. Основы информационной безопасности: учебное пособие / Иркутск: ФГОУ ВПО ВСИ МВД России. - 2015. - 98 с.

[3] Кузнецов, А.И. Двоичная тайнопись (по материалам открытой печати) / КомпьютерПресс - 2004. № 4. - с. 38-41.

[4] Аненко, М.Н., Пархоменко, И.И. Способы и методы осуществления атак на системы цифровых водяных знаков [Электронный ресурс] / URL: http://www.rusnauka.com/35_OINBG_2012/Informatica/4_122942.doc.htm (дата обращения 20.12.2020). - Загл. с экрана. - Яз. рус.

защиты информации. В отличие от криптографии, основной задачей стеганографии является сокрытие самого факта передачи сообщения[5].

В связи с развитием и распространением компьютерных технологий, в качестве контейнера применяются разнообразные данные: текстовые документы, аудиофайлы, изображения, видеофайлы и многие другие. Значительный размер медиафайлов, их информационная избыточность, а также возросшая популярность передачи их через сеть делает аудио, видео и графические файлы особенно привлекательными для использования в качестве стегоконтейнера.

Цифровая стеганография разделяется на следующие направления[6]:

- встраивание ЦВЗ;
- встраивание заголовков;
- встраивание информации с целью ее скрытой передачи;
- встраивание идентификационных номеров.

Технология ЦВЗ создана для защиты авторских прав мультимедийных файлов. Зачастую ЦВЗ невидимы. Однако они могут быть видимыми на изображении или видео. Обычно ЦВЗ представляют собой текст или логотип, идентифицирующий автора. Но, к сожалению, ЦВЗ, как и криптоалгоритмы, имеют свои недостатки, поэтому перед тем как выбрать тот или иной алгоритм формирования ЦВЗ необходимо разобраться, что может предпринять злоумышленник для того, чтобы избавится от ЦВЗ.

Из ранее сказанного можно сделать вывод, что маркирование ЦВЗ является важной мерой обеспечения безопасности для защиты авторского права. Объемы создаваемой и защищаемой информации постоянно растут, вместе с тем развиваются и методы атак со стороны злоумышленников. Поэтому исследования в данной области будут актуальны еще долгое время. Рассмотрение существующих методов борьбы с известными атаками помогает

---

[5] Кокорин, П.П. О методах стегоанализа в аудиофайлах / Труды СПИИРАН. Вып. 4. - СПБ.: Наука. - 2007. - с. 239-246.

[6] Грибунин, В.Г., Оков, И.Н., Туринцев, И.В. Цифровая стеганография: Стратегия развития информационного общества в РФ. / М.: Солон-Пресс. - 2009. - 265 с.

выявлять наиболее уязвимые места при создании ЦВЗ и совершенствовать методы их внедрения. Изучение возможных атак помогает выявить те, к которым создаваемые ЦВЗ наиболее уязвимы, а также разрабатывать новые методы повышения устойчивости ЦВЗ[7].

Целью данной работы является исследование ЦВЗ встроенных методом DEW на устойчивость к определенным видам атак.

Для достижения этой цели поставлены следующие задачи:

- рассмотреть типы ЦВЗ;
- изучить алгоритм встраивания ЦВЗ методом DEW;
- рассмотреть методы оценки качества изображений;
- изучить методы атак на системы ЦВЗ;
- разработать и реализовать программное обеспечение, позволяющее производить атаки на ЦВЗ различных типов;
- протестировать работоспособность разработанного программного обеспечения;
- с помощью разработанного программного обеспечения произвести атаки на базу изображений с внедренными методом DEW цифровыми водяными знаками;
- проанализировать данные, полученные с помощью разработанного программного обеспечения;
- на основе полученных и проанализированных данных оценить стойкость ЦВЗ встроенных методом DEW к атакам.

Дипломная работа состоит из введения, 5 разделов, заключения, списка использованных источников и 5 приложений. Общий объем работы – 65 страниц, из них 49 страниц – основное содержание, включая 43 рисунка и 2 таблицы, список использованных источников из 35 наименований.

[7] Петелина, М.В. Устойчивость цифровых водяных знаков / Сборник трудов молодых ученых и сотрудников кафедры ВТ / под ред. д.т.н., проф. Т.И. Алиева. - СПб: СПбНИУ ИТМО. - 2012. - с. 66-69. - 94 с.

КРАТКОЕ СОДЕРЖАНИЕ

В первом разделе рассматривается технология цифровых водяных знаков, их типы и области применения. Особое внимание уделено прозрачным цифровым водяным знакам.

Обычно цифровые водяные знаки классифицируются по следующим основным параметрам:

- объем;
- сложность;
- обратимость;
- прозрачность;
- надежность;
- безопасность;
- верификация.

Во втором разделе рассматривается алгоритм внедрения цифрового водяного знака Differential energy watermark (DEW), предложенный Лангелааром. Данный метод основан на выборочном отбрасывании части высокочастотных коэффициентов дискретного косинусного преобразования (ДКП) сжатых изображений[8].

Метод DEW осуществляет внедрение цифрового водяного знака, состоящего из l бит $b_j(j = 0, 1, 2, \dots, l-1)$ в JPEG-изображения. Каждый бит цифрового водяного знака встраивается в выбранную область, состоящую из n блоков по $8x8$ коэффициентов дискретного косинусного преобразования канала яркости изображения каждый[9].

Каждый бит цифрового водяного знака внедряется в выбранную область модификацией разности энергий D между высокочастотными коэффициентами ДКП верхней части этой области (субобласть A) и ее нижней части (субобласть B). Подмножество высокочастотных коэффициентов обозначается $S(c)$.

---

[8] Иваненко, В.Г., Ушаков, Н.В. Защита изображений формата JPEG при помощи цифровых водяных знаков / Безопасность информационных технологий. - 2018. - № 2. - с. 106-113.

[9] Hanjalic, A., Langelaar, G.C., Roosmalen, P.M.B., Biemond, J., Lagendijk, R.L. Image and Video Databases: Restoration, Watermarking and Retrieval / Elsevier Science B.V. - 2000. - vol. 8. - p.468.

Энергия субобласти A вычисляется по формуле (1):

$$E_A(c,n,Q) = \sum_{d=0}^{\frac{n}{2}-1} \sum_{i \epsilon S(c)} (\left[\theta_{i,d}\right]_Q)^2, \quad (1)$$

где $\Theta_{i,d}$ – коэффициент ДКП с индексом i из d-го блока коэффициентов ДКП субобласти A;

$[\,]_Q$ – означает, что энергия вычисляется у квантованных коэффициентов[10].

Энергия субобласти B вычисляется аналогичным способом (2):

$$E_B(c,n,Q) = \sum_{d=\frac{n}{2}}^{n-1} \sum_{i \epsilon S(c)} (\left[\theta_{i,d}\right]_Q)^2. \quad (2)$$

Подмножество $S(c)$ определяется на основе выбранного порога, рассчитываемого по формуле (3):

$$S(c) = \{h \epsilon \{1,63\} | (h \geq c)\}. \quad (3)$$

Выбор подходящего значения порога крайне важен, так как этим определяется стойкость ЦВЗ к удалению и его заметность на изображении.

Преимущества алгоритма DEW:

- Вносит в изображение меньше искажений, чем иные методы встраивания ЦВЗ.
- Для удаления цифрового водяного знака требуется проведение вычислительных операций, более сложных, чем встраивание нового произвольного водяного знака.

Недостатки алгоритма DEW:

- Высокочастотные коэффициенты ДКП легко отбрасываются фильтрами, в связи с чем алгоритм будет уязвим к этому воздействию на контейнер.
- Алгоритм не учитывает, какое влияние на исходное изображение оказывает отбрасывание коэффициентов ДКП.

В третьем разделе рассматривается вопрос количественной оценки качества изображения. Существующие методы, также именуются метриками, оценки можно разделить на субъективные и объективные.

---

[10] Langelaar, G.C., Lagendijk, R.L. Optimal differential energy watermarking of DCT encoded images and video / IEEE Transactions on Image processing. - 2005. - vol. 10, no. 1. - pp. 148-158.

Субъективные метрики основаны на перцептивной оценке человеком-наблюдателем, в то время как объективные метрики основаны на вычислительных моделях, которые пытаются оценить качество изображения. Субъективные метрики часто являются более точными с точки зрения восприятия, однако не все из них удобны: некоторые из них трудоемки или дороги для вычисления. Также результаты оценок субъективным и объективным способами часто могут не совпадать. По этой причине для анализов результатов предпочтительно использовать метрики из обеих категорий[11].

В рамках данной работы кратко рассматриваются такие из наиболее часто используемых метрик, как: пиковое отношение сигнала к шуму и структурное сходство.

Пиковое отношение сигнала к шуму (PSNR) – это наиболее часто используемая объективная метрика. Ее значение можно определить, как величину, обратно пропорциональную логарифму средней квадратичной ошибки между изображением-оригиналом и измененной копией[12].

Структурное сходство (SSIM) – это субъективная метрика, используемая для оценки структурного сходства между изображениями на основе сравнения яркости, контрастности и структуры[13].

Поскольку метрика структурного сходства оценивает качество реконструкции с точки зрения зрительной системы человека, она лучше соответствует требованиям перцептивной оценки.

В четвертом разделе рассматриваются виды атак на цифровые водяные знаки. Успешная атака должна повредить или уничтожить цифровой водяной знак при сохранении качества изображения. Для уничтожения следов присутствия цифровых водяных знаков могут быть использованы как определенные виды атак отдельно, так и их совокупность.

---

[11] Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P. Image Quality Assessment: From Error Visibility to Structural Similarity / IEEE Transactions on Image Processing. - 2004. - vol. 13, no. 4. - pp. 600-612.

[12] Wang, Z., Chen, J., Hoi, S.C.H. Deep Learning for Image Super-resolution: A Survey / IEEE Transactions on Patter Analysis and Machine Intelligence. - 2020.

[13] Старовойтов, В.В. Уточнение индекса SSIM структурного сходства изображений / Информатика. - 2018. - т. 15, № 3. - с. 41-55.

К группе атак, направленных на удаление ЦВЗ, относятся такие атаки, как: шумоподавление, перемодуляция, сжатие с потерями, усреднение и атака сговором.

Геометрические атаки предназначены не для удаления самого встроенного ЦВЗ, а для его искажения посредством пространственных изменений стегоданных. Геометрические атаки математически моделируются как аффинные преобразования с неизвестным декодеру параметром. Всего имеется шесть аффинных преобразований: масштабирование, изменение пропорций, повороты, сдвиг и усечение.

Криптографические атаки названы так потому, что они имеют аналоги в криптографии. Данный вид атак основан на принципе создания ложного ЦВЗ. К ним относятся атаки с использованием оракула, а также взлома при помощи «грубой силы»[14].

Атаки против используемого протокола служат для воздействия на концепцию приложения для создания ЦВЗ. Идея состоит в том, что злоумышленник, имеющий копию стегоданных, может утверждать, что данные также содержат ЦВЗ злоумышленника, вычитая его собственный ЦВЗ. Это может создать ситуацию двусмысленности в отношении реального владения данными[15].

В пятом разделе приведены сведения об использованных средствах разработки, о разработанном программном обеспечении и его эффективности, а также анализ полученных в ходе работы данных.

При разработке программного обеспечения в ходе выполнения работы был использован язык программирования Python (версия Python 3). Для создания графического интерфейса пользователя была использована графическая библиотека Tkinter.

---

[14] Cox, I.J., Linnartz, J.-P.M.G. Some general methods for tampering with watermarks / IEEE Journal on Selected Areas in Communications. - 1998. - vol. 16, no. 4. - pp. 587-593.

[15] Kutter, M., Voloshynovskiy, S., Herrigel, A. Watermark copy attack / P.W. Wong, E.J. Delp (Eds.) / IS&T/SPIE's 12th Annual Symposium, Electronic Imaging 2000: Security and Watermarking of Multimedia Content II. - 2000. - vol. 3971. - pp. 23-28.

Разработанное ПО реализует следующий функционал:

1) произвести атаки на выбранное изображение;
2) проанализировать качество полученного после атаки изображения (методы PSNR и SSIM);
3) оценить эффективность встроенного ЦВЗ к атаке;
4) просмотреть изображение.

Для тестирования программного обеспечения использовалось более 100 различных изображений с внедренным цифровым водяным знаком.

Результаты, полученные в процессе тестирования, были проанализированы и обобщены.

В таблице 1 приведены обобщенные данные по результатам атак на ЦВЗ, внедренные методом DEW. Для анализа была использована база, содержащая более 100 различных изображений, с внедренным ЦВЗ.

Таблица 1 - Обобщенные данные по результатам атак на ЦВЗ

Тип атаки	**Коэффициент Пирсона**	**SSIM**
Обрезка	≈0.5	–
Масштабирование	<0.1	–
Поворот	≈0.5	–
Сжатие	>=0.5	>0.7
Шумоподавление	>0.3	>0.75
Внесение шума	>0.5	≈0.5
Внедрение нового ЦВЗ	>0.5	>0.7

Опираясь на полученные результаты, можно сделать вывод, что метод DEW можно считать устойчивым ко многим видам воздействий на контейнер.

ЗАКЛЮЧЕНИЕ

В процессе написания данной работы были рассмотрены технология цифровых водяных знаков, типы ЦВЗ и области их применения. Был подробно изучен алгоритм встраивания ЦВЗ методом DEW. Также были рассмотрены методы оценки качества изображений. Особое внимание было уделено методу пикового отношения сигнала к шуму и методу структурного подобия. Были изучены методы атак на цифровые водяные знаки. Подробно рассмотрены такие типы атак, как: атаки, направленные на удаление ЦВЗ; геометрические атаки; криптографические атаки и атаки против используемого протокола.

В практической части было разработано и реализовано программное обеспечение с простым графическим интерфейсом, содержащее модули для проведения тестирования различных алгоритмов встраивания ЦВЗ на устойчивость к определенным видам атак, для оценки эффективности ЦВЗ и для оценки перцептивного качества полученного в результате атаки изображения. Тестирование программного обеспечения дало положительные результаты.

Также в ходе практической части работы с помощью разработанного программного обеспечения были произведены атаки на базу, содержащую более 100 изображений, с ЦВЗ встроенными методом DEW. Полученные данные были проанализированы и обобщены. Для оценки устойчивости цифрового водяного знака применялись коэффициент Пирсона и подсчет процента побитового совпадения между внедренным и извлеченным ЦВЗ.

В процессе анализа устойчивости ЦВЗ встроенных методом DEW к атакам были получены следующие результаты:

- Геометрические атаки:
  - Обрезка – коэффициент Пирсона ≈0.5
  - Масштабирование – коэффициент Пирсона <0.1
  - Поворот – коэффициент Пирсона ≈0.5
- Статистические атаки:
  - Сжатие – коэффициент Пирсона >=0.5
  - Шумоподавление – коэффициент Пирсона >0.3

- Внесение шума – коэффициент Пирсона >0.5
- Внедрение нового ЦВЗ – коэффициент Пирсона >0.5

Опираясь на полученные результаты, был сделан вывод, что метод DEW можно считать в достаточной мере устойчивым ко многим видам воздействий на контейнер.

Результатом данной работы является полученная оценка эффективности ЦВЗ, внедренного методом DEW, а также разработанное ПО, позволяющее производить атаки на ЦВЗ различных типов. На основе полученных результатов существует возможность дальнейшей модификации алгоритма встраивания ЦВЗ методом DEW для достижения большей устойчивости к таким атакам, как: шумоподавление и масштабирование. Кроме того, существует возможность дальнейшей модификации разработанного программного обеспечения с целью расширения функционала.

В данной работе все поставленные цели были достигнуты и все поставленные задачи выполнены.