ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ

УДК 004.056

***А.Е. Лагун**, канд. техн. наук, доцент*
(Львівський державний університет безпеки життєдіяльності)

АТАКИ НА СУЧАСНІ СТЕГАНОГРАФІЧНІ СИСТЕМИ І МЕТОДИ ЗАХИСТУ

Проведено дослідження атак на стеганографічні системи, зокрема системи цифрових водяних знаків. Здійснено класифікацію атак на основі заповненого контейнера. Особливістю таких атак є те, що вони не змінюють контейнера з прихованою інформацією. Крім того, проаналізовано атаки, які мають аналоги в криптоаналізі, а саме атаки на основі відомого заповненого контейнера або відомого вбудованого повідомлення, вибраного заповненого контейнера або вибраного вбудованого повідомлення.

Також розглянуто особливості атак на стеганографічні системи інтелектуальної власності, метою яких є руйнування або видалення цифрового водяного знака. Зокрема проаналізовано і наведено рекомендації із використання методів захисту від геометричних, криптографічних атак, а також атак на видалення цифрового водяного знака і проти використаного для вбудовування цифрового водяного знака протоколу.

Ключові слова: стеганографія, стеганографічна система, стеганоаналіз, атака, цифровий водяний знак, порожній і заповнений контейнер, порушник, стійкість, приховане повідомлення.

А.Е. Лагун

АТАКИ НА СОВРЕМЕННЫЕ СТЕГАНОГРАФИЧЕСКИЕ СИСТЕМЫ И МЕТОДЫ ЗАЩИТЫ

Проведено исследование атак на стеганографические системы, в частности системы цифровых водяных знаков. Осуществлена классификация атак на основе заполненного контейнера. Особенностью таких атак является то, что они не изменяют контейнера со скрытой информацией. Кроме того, проанализированы атаки, которые имеют аналоги в криптоанализе, а именно атаки на основе известного заполненного контейнера или известного встроенного сообщения, избранного заполненного контейнера или избранного встроенного сообщения.

Также рассмотрены особенности атак на стеганографические системы интеллектуальной собственности, целью которых является разрушение или удаление цифрового водяного знака. В частности проанализированы и приведены рекомендации по использованию методов защиты от геометрических, криптографических атак, а также атак на удаление цифрового водяного знака и против использованного для встраивания цифрового водяного знака протокола.

Ключевые слова: стеганография, стеганографическая система, стегоанализ, атака, цифровой водяной знак, пустой и заполненный контейнер, нарушитель, устойчивость, скрытое сообщение.

A.E. Lagun

ATTACKS ON MODERN STEGANOGRAPHIC SYSTEMS AND SECURITY METHODS

In the article attacks on steganographic systems (including digital watermarking systems) have been investigated. Classification of attacks based on the filled container has been suggested. These attacks do not change the container with hidden information and this is their main peculiarity. In particular, attacks that have analogs in cryptanalysis have been analyzed. These are attacks based on known filled container or embedded message and also selected filled container or selected embedded message.

Attacks on steganographic systems of the intellectual property, whose purpose is the destruction or removal of a digital watermark have been also analized in the article. Methods of protection against geometric attacks, cryptographic attacks, attacks on removing digital watermark and attacks against protocol, used for embedding a digital watermark have been sugested.

Key words: steganography, steganography system, steganalysis, attack, digital watermark, empty and filled container, intruder, stability, hidden message.

Вступ

Слово "стеганографія" в перекладі з грецької означає "таємнопис" Він може реалізуватися різними способами. В класичному випадку приховуване повідомлення вбудовується в деякий фізичний об'єкт, який не привертає уваги. Потім цей об'єкт відкрито передається адресату.

На сьогодні через використання комп'ютерів в усіх життєвих сферах розвинулася комп'ютерна стеганографія, яка в основному використовує цифрову обробку сигналів. В [3] визначено такі напрями комп'ютерної стеганографії:

1) вбудовування інформації у фізичні об'єкти з метою її прихованої передачі для забезпечення конфіденційності;

2) вбудовування цифрових водяних знаків (ЦВЗ) з метою підтвердження достовірності;

3) вбудовування ідентифікаційних номерів та заголовків.

Фізичний об'єкт, в який вбудовується прихована інформація, називається контейнером. Контейнер може бути заповнений і незаповнений. Як правило, контейнером є цифровий носій інформації, що має аналогову природу – аудіофайли, відеофайли та нерухомі зображення.

Завдання вбудовування і видобування інформації з контейнера виконує стеганосистема, яка складається з стеганокодера і стеганодекодера. Стеганокодер перетворює приховане повідомлення до вигляду, зручного для вбудовування в сигнал-контейнер і вбудовує приховане повідомлення в сигнал-контейнер з урахуванням його моделі. Стеганодекодер визначає наявність прихованого повідомлення в контейнері і, за наявності, видобуває і відновлює приховане повідомлення.

Стеганографія тісно пов'язана з криптографією, проте ці науки мають різні підходи до захисту інформації. Зокрема криптографія приховує інформацію за допомогою операції шифрування, тобто наперед відомо, що в криптограмі міститься зашифрована інформація. В свою чергу стеганографія приховує факт наявності секретної інформації, тому заповнений контейнер не повинен відрізнятися від порожнього. Для підвищення захищеності інформації методи криптографії і стеганографії можуть поєднуватися.

Стеганосистема утворює стеганоканал, по якому передається заповнений контейнер. Доступ до цього каналу можуть отримати порушники. Опишемо коротко, якої шкоди можуть завдати порушники. Для таємного обміну повідомленнями двоє адресатів повинні мати відомий обом секретний ключ, який визначатиме місцезнаходження прихованого повідомлення.

В першу чергу порушник може встановити факт наявності стеганоканалу і читати повідомлення. Можливість читання повідомлення визначається стійкістю використаної системи приховування. Розглянутий тип порушників вважається пасивним.

Існує також активний порушник, який може видаляти або руйнувати приховані повідомлення. Хоча факт втручання буде відомий, проте мета порушника – зламування стеганосистеми – буде досягнута. Найбільш небезпечним є зловмисний порушник, який, крім руйнування, може здійснювати підміну стеганоповідомлень.

Загальна класифікація атак на стеганосистеми

Для реалізації загроз порушники використовують атаки. Метою роботи є аналіз та класифікація атак, що можливі в стеганосистемах. Спочатку розглянемо загальні типи атак, що застосовуються до стеганосистем.

Найпростішою з атак є суб'єктивна, яка полягає у визначенні за допомогою психовізуального спостереження та найпростішого аналізу заповненого контейнера наявності прихованого

повідомлення у стегоконтейнері. Дана атака, як правило, здійснюється на першому етапі розкриття стеганосистеми і є ефективною лише в повністю незахищених стеганосистемах.

Аналіз заповненого контейнера передбачає проведення таких дій [2]:

– виявлення прихованого повідомлення за зовнішніми ознаками;

– спроба використати відомий алгоритм вбудовування до заповненого контейнера;

– здійснення аналізу для окремих ділянок контейнера або порівняння кількох заповнених контейнерів;

– виділення прихованого повідомлення за відомим алгоритмом вбудовування, проте невідомим ключем.

Проаналізуємо атаки, які визначаються властивостями використаного в стеганосистемі контейнера для передавання прихованого повідомлення. Особливістю цих атак є те, що вони не змінюють заповненого контейнера.

Найпростішою є атака на основі відомого порожнього контейнера, яка дає змогу встановити наявність стеганоканалу, якщо порівняти порожній і заповнений контейнери. Можлива також атака з обраним порожнім контейнером, при якій порушник нав'язує адресатам свій контейнер з властивостями, що зменшують секретність вбудовування. Наприклад, контейнер у вигляді однотонного зображення дозволяє легко виявити факт наявності прихованого повідомлення. Ще одна атака використовує відому математичну модель контейнера. В цьому випадку порушник виявляє відмінність між відомою моделлю контейнера і моделлю заповненого контейнера. Протидія цій атаці полягає у вбудовуванні повідомлення із збереженням статистики порожнього і заповненого контейнерів.

Як було зазначено вище, стеганографія тісно пов'язана з криптографією, тому в стеганосистемах можливі криптографічні атаки.

Аналогічно до криптоаналізу, в стеганоаналізі бувають такі атаки [4, 10]:

– атака на основі відомого вбудованого повідомлення відбувається в системах захисту інтелектуальної власності, коли потрібно визначити цифровий водяний знак; при отриманні ключа місцезнаходження ЦВЗ буде виявлено і тому стеганосистему буде зламано;

– в атаці на основі відомого заповненого контейнера передбачається, що порушник має один або декілька заповнених контейнерів, а вбудовування прихованої інформації у всі контейнери відбувалося за однаковим алгоритмом; тоді порушник повинен виявити стеганоканал і визначити ключ, що дасть змогу йому розкрити приховане повідомлення;

– атака на основі вибраного заповненого контейнера використовується в системах захисту інтелектуальної власності для компрометації ЦВЗ; аналізуючи результати декодування різних заповнених контейнерів, порушник намагається розкрити ключ;

– в атаці на основі вибраного прихованого повідомлення порушник надає учасникам обміну інформацією свої повідомлення і, проаналізувавши одержані заповнені контейнери, може виявити приховане повідомлення;

– адаптивна атака на основі вибраного прихованого повідомлення подібна до попередньої, а саме, порушник нав’язує свої повідомлення учасникам обміну повідомленнями адаптивно, залежно від результатів аналізу попередніх заповнених контейнерів.

Класифікація атак на стеганосистеми інтелектуальної власності

Вважаємо, що як контейнер використовується цифрове зображення у вигляді матриці пікселів. Розглянемо найпростіший випадок півтонового цифрового зображення в градаціях сірого, в якому кожен піксел кодується вісьмома бітами. Відомо, що людське око не здатне помітити зміну зображення внаслідок модифікації молодшого значущого біта [9]. Така властивість використовується в алгоритмах стиснення інформації, а також в стеганографії. Таким чином, для вибраного зображення максимальний обсяг вбудованих даних може становити 1/8 від обсягу контейнера, зокрема в зображення розміром 1024 х 1024 можна вбудувати приблизно 130 кілобайт інформації. Зрозуміло, що при використанні кольорових зображень для приховування використовуються молодші значущі біти з RGB-палітри.

Розглянутий алгоритм вбудовування, попри свою простоту, є непрактичним у разі застосування до зображення з прихованою інформацією різних способів обробки зображень

[8]. Наприклад, при очищенні сигналів від шумів молодший значущий біт в більшості випадків буде сприйматися як шум, тому він буде вилучений із зображення; алгоритми стиснення також використовують значення молодшого значущого біта для зменшення розміру зображення, тому цим значенням можна знехтувати при стисненні зображення.

В [1] запропоновано таку класифікацію атак на стеганосистеми інтелектуальної власності:

- атаки, метою яких є видалення цифрового водяного знака;
- геометричні атаки, що спотворюють заповнений контейнер;
- атаки проти використаного протоколу вбудовування і перевірки цифрового водяного знака;
- криптографічні атаки.

Атаки, які видаляють цифрові водяні знаки

Цей тип атак використовує характеристику цифрового водяного знака у вигляді статистично описаного шуму. Очищення від шуму полягає у фільтрації сигналу з використанням різних статистичних критеріїв. Стиснення з втратами і очищення сигналів від шумів зменшують пропускну здатність каналу, особливо при наявності однорідних областей зображення, коефіцієнти перетворення яких можуть бути обнулені без помітного зниження якості відновленого зображення.

Існує ефективна для високочастотного цифрового водяного знака атака перемодуляції, яка робить спробу обману декодера при виявленні ЦВЗ. Останній передбачається шляхом порівняння фільтрованої версії зображення і заповненого контейнера, тому реальні ЦВЗ будуються так, щоб їх спектр відповідав спектру початкового зображення. Після віднімання високочастотної частини ЦВЗ низькочастотна залишається незмінною, що свідчить про наявність ЦВЗ в зображенні. В свою чергу, високочастотна складова скомпенсує низькочастотну і ЦВЗ не буде знайдений. Протидією цій атаці є виконання низькочастотної фільтрації.

Ефективними також є атаки видалення цифрових водяних знаків, які пердбачають наявність великої кількості заповнених контейнерів з різними ЦВЗ або з різними ключами. Зокрема, при атаці усереднення атакуючий може одержати узагальнений цифровий водяний знак і відняти його від зображення. Атака змови передбачає розбиття різних заповнених контейнерів на частини, з яких створюється множина, на яку здійснюється атака. Із збільшенням кількості заповнених контейнерів у порушника збільшується можливість виявити ЦВЗ. Захистом від атаки змови є спеціальна побудова заповненого контейнера. Ще одна ефективна атака на ЦВЗ називається мозаїчною. При ній зображення розбивається на декілька частин таким чином, що цифровий водяний знак неможливо знайти.

Геометричні атаки проти заповненого контейнера

Метою геометричних атак є зміна цифрового водяного знака шляхом внесення спотворень у заповнений контейнер. Як правило, геометричні атаки використовують такі афінні перетворення, як повороти, зсув, обрізання, зміна пропорцій і масштабування. Такі атаки можуть застосовуватися як до цілого зображення, так і до його частин, зокрема обертають окремі частини зображення, вирізають окремі пікселі або рядки зображення, помінявши їх місцями та інше.

При геометричних атаках можливе усунення синхронізації ЦВЗ. Якщо визначити метод синхронізації, то його можна зруйнувати шляхом згладжування максимумів в амплітудному спектрі цифрового водяного знака. Для забезпечення синхронізації максимуми вбудовуються в спектральній області, або послідовність цифрових водяних знаків повторюється через певний період. Розглянута геометрична атака руйнує максимуми, усуваючи синхронізацію.

Сучасні методи вбудовування ЦВЗ стійкі до атак на ціле цифрове зображення завдяки використанню спеціальних методів відновлення синхронізації. Проте, забезпечення стійкості до атак на окремі частини зображення є актуальною проблемою через особливість системи людського зору.

Атаки проти використаного протоколу вбудовування

Розглянемо протокол. Порушник перехоплює заповнений контейнер і визначає деяку частину захищеного зображення як цифровий водяний знак. Потім він створює інший контейнер, віднявши цю частину зображення, і відправляє цей помилковий контейнер стеганодекодеру.

Атака такого типу називається інверсною. І в помилковому контейнері існує справжній цифровий водяний знак, і в оригінальному заповненому контейнері існує оголошений порушником помилковий ЦВЗ, що робить неможливим процес виявлення ЦВЗ. Якщо у стеганодекодера є оригінальне зображення, то інтелектуальну власність буде підтверджено. Захистом від цієї атаки є створення залежності цифрового водяного знака від зображення за допомогою односторонньої функції.

В комп'ютерних системах при поширенні ліцензійного програмного забезпечення використовується захист за допомогою цифрових водяних знаків, який полягає в зміні властивостей ЦВЗ після копіювання для запобігання повторному копіюванню. Завданням порушника є створити копію, яка дуже близька до оригіналу. Тоді він, маючи доступ до повідомлень до і після вбудовування ЦВЗ, може обчислити різницю між цими повідомленнями. Віднявши цю різницю від початкового повідомлення, порушник отримає якесь спотворене повідомлення. На останньому кроці він формує копію, додавши до спотвореного зображення різницю між спотвореним та оригінальним повідомленням з ЦВЗ. Остання копія майже не відрізняється від оригіналу. Якщо цифровий водяний знак не буде стійким до адитивного шуму, яким і буде остання різниця, то мета порушника буде досягнута.

Криптографічні атаки в стеганоаналізі

При зламуванні стеганосистем можуть використовуватися атаки методом перебору і з використанням оракула. Як і для криптоаналізу, одержання ключа методом перебору є неефективним і може використовуватися як доповнення до інших відомих атак. Розглянемо атаку з використанням оракула, метою якої є видалення цифрового водяного знака [1].

В цій атаці передбачається наявність у порушника детектора, що може формувати повідомлення про наявність чи відсутність ЦВЗ в зображенні. На початковому етапі атакуючий подає на вхід детектора різні зображення, вивчаючи його поведінку. Модифікуючи попіксельно зображення порушник може з'ясувати алгоритм роботи детектора.

Можлива інша атака, яка дає змогу порушнику вилучити ЦВЗ із зображення.

Крок 1. Маючи початкове зображення з цифровим водяним знаком, порушник створює модифіковане зображення, змінюючи пікселі початкового, доки детектор не покаже відсутності ЦВЗ.

Крок 2. Порушник збільшує або зменшує значення кожного пікселя модифікованого зображення, доки детектор не покаже наявність ЦВЗ знову, щоб виявити збільшення чи зменшення значення кожного пікселя ЦВЗ.

Крок 3. Порушник визначає ті пікселі зображення, модифікація яких порушить роботу детектора, проте майже не вплине на якість зображення, і віднімає ці пікселі від початкового зображення.

Висновки і рекомендації щодо захисту стеганосистем з використанням цифрових водяних знаків

Розглянемо докладніше методи захисту від наведених вище атак на системи цифрових водяних знаків.

Одним із способів захисту від атак, що використовують афінні перетворення, є використання блокового детектора. Зображення, що містить ЦВЗ, розбивають на блоки, а потім до кожного із блоків застосовують різні афінні перетворення, кожен раз фіксуючи відмінності між початковим блоком і зміненим. Перетворення, які найбільше змінюватимуть блоки, вважаються такими, які здійснив атакуючий. Тоді, застосувавши зворотні перетворення до виявлених, можна отримати початкове зображення.

Також геометричним афінним атакам можна запобігти, якщо вбудовувати ЦВЗ у візуально помітні частини зображення, які не можна видалити без спотворення зображення. Мо-

жливе також використання додаткового цифрового водяного знака, який фіксуватиме дії порушника, проте основною вимогою в цьому випадку є атака саме на додатковий ЦВЗ.

Протидією атакам на виявлення і видалення цифрового водяного знака є адаптація спектра ЦВЗ до спектра сигналу-контейнера. Оскільки спектральна густина ЦВЗ значно менша за спектральну густину контейнера, то виявити такий ЦВЗ надзвичайно важко. На практиці адаптація спектра ЦВЗ можлива лише за наявності контейнера.

Для захисту інтелектуальної власності в стеганосистемах можна використати класичний для криптографії протокол з мітками часу [6-7]. А саме в цифровий водяний знак вбудовується мітка часу. Особа, яка буде мати в зображенні більш ранню мітку часу, вважатиметься справжнім власником. Інший спосіб передбачає те, що ЦВЗ повинен бути адаптивним до сигналу і вбудовуватися за допомогою односторонньої хеш-функції. Ця функція перетворить біти зображення в послідовність бітів, а потім, залежно від значення біту, використовуються дві функції вбудовування ЦВЗ.

Для захисту від криптографічних атак основну увагу потрібно приділити ключовій псевдовипадковій послідовності, згідно з якою вбудовується прихована інформація. Правильний вибір параметрів псевдовипадкової послідовності (перш за все вона повинна бути криптографічно стійкою) може значно підвищити стійкість стеганосистем до атак додавання шуму, стиснення та інших.

Метою подальших досліджень є моделювання описаних стеганографічних атак в реальній комп'ютерній мережі і виявлення нових способів протидії цим атакам.

Список літератури:

1. Грибунин В. Г. Цифровая стеганография / В. Г. Грибунин, И. Н. Оков, И. В. Туринцев. – М. : СОЛОН-Пресс, 2002.
2. Хорошко В. О. Основи комп’ютерної стеганографії : навчальний посібник для студентів і аспірантів / В. О. Хорошко, О. Д. Азаров, М. Є. Шелест, Ю. Є. Яремчук. – Вінниця : ВДТУ, 2003.
3. Конахович Г. Ф. Компьютерная стеганография / Г. Ф. Конахович, А. Ю. Пузыренко. – К. : МК-Пресс, 2006. – 288 с.
4. Аграновский А. В. Стеганография, цифровые водяные знаки и стегоанализ [Текст] / А. В. Аграновский, А. В. Балакин, В. Г. Грибунин. – М. : Вузовская книга, 2009. – 220 с.
5. Cox J. Watermarking as Communications With Side Information / Cox J., Miller M., McKellips A. // Proceedings of the IEEE. – 1999. Vol. 87. № 7. P. 1127 – 1141.
6. Zhao J. Embedding Robust Labels into Images for Copyright Protection / J. Zhao, E. Koch // Proceeding of the Int. Congress on Intellectual Property Rights for Specialized Information, Knowledge and New Techniques. – Munich – Vienna, Verlag, Aug. 1995. – Pp. 242 – 251.
7. Koch E. Towards Robust and Hidden Image Copyright Labeling / E. Koch, J. Zhao // IEEE Workshop on Nonlinear Signal and Image Processing. – Greece, June 20 – 22, 1995. Pp.123 – 132.
8. Кущ С. М. Виявлення прихованих повідомлень як складова комплексних систем захисту інформації / С. М. Кущ, В. М. Луценко, Д. О. Прогонов // Захист інформації. – 2012. № 3. – С. 65 – 71.
9. Корольов В. Ю. RS- стеганоаналіз. Принципи роботи, недоліки та концепція методу його обходу / В. Ю. Корольов, В. В. Поліновський, В. А. Герасименко // Вісник Вінницького політехнічного інституту. – 2010. № 6. – С. 66 – 71.
10. Кошкина Н. В. Обзор и классификация методов стеганоанализа / Н. В. Кошкина // УСиМ. – 2015. № 3. – С. 3 – 12.

References:

1. Gribunin, V. G., Okov, I. N., Turintsev, I. V. (2002). *Digital steganography.* Moscow: SOLON-PRESS (in Russ.)

2. Khoroshko, V. O., Azarov, O. D., Shelest, M. Ye, Yaremchuk, Yu. Ye. (2003). *Basics of computer steganography*. – Vinnytsya: VSTU (in Ukr.)

3. Konakhovych, G. F., Puzyrenko, A. Yu. (2006). *Computer steganography.* – Kyiv: MK-PRESS (in Russ.)

4. Agranovskij, A. V., Balakin, A. V., Gribunin, V. G. (2009). *Steganography, digital watermarking and steganalysis* (Text). – Moscow: The university book (in Russ.)

5. Cox, J., Miller, M., McKellips, A. (1999). *Watermarking as Communications With Side Information.* (Proceedings of the IEEE, Vol. 87б), 7, 1127 – 1141.

6. Zhao, J., Koch, E. (1995). *Embedding Robust Labels into Images for Copyright Protection.* (Proceeding of the Int. Congress on Intellectual Property Rights for Specialized Information, Knowledge and New Techniques). Munich - Vienna, Verlag, 242 – 251.

7. Koch, E., Zhao, J. (1995). *Towards Robust and Hidden Image Copyright Labeling.* (IEEE Workshop on Nonlinear Signal and Image Processing). Greece, 123 – 132.

8. Kushch, S. M., Lutsenko, V. M., Progonov, D. O. (2012). *Detection of hidden messages as part of complex information security systems.* – (Information Security), 3, 65 – 71 (in Ukr.)

9. Korolov, V. Yu., Polinovskyj, V. V., Gerasymenko, V. A. (2010). RS- *steganalysis. Principles of work, shortcomings and concept of circumvention method.* – (Visnyk VPI), 6, 66 – 71 (in Ukr.)

10. Koshkina, N. V. (2015). *Overview and classification of steganalysis methods.* – (CS&M). 3, 3 – 12. (in Russ.)

❖❖❖