УДК 004.056.53

О.В. Нарожний,
к.т.н., доцент
Одеський національний політехнічний університет
e-mail: spawns7650@ukr.net

В.В. Коленко,
аспірант,
Одеський національний політехнічний університет
e-mail: violka86@mail.ru

АНАЛІЗ АТАК ТА МЕТОДИ ЗАХИСТУ СТЕГОСИСТЕМ

В.В. Коленко, О.В. Нарожний. Аналіз атак та методи захисту стегосистем. Розглянуто існуючі види атак на стегосистеми з цифровим водяним знаком. Запропоновано використання мультиагентних технологій для вирішення проблеми захисту даних стегосистем.

V.V. Kolenko, A.V. Narozhnij. Analysis of attacks and defence methods of steganosystems. The existent types of attacks are considered on steganosystems with a digital thread-mark. The use of multi-agent technologies is offered for the decision of problem of data protection for steganosystems.

Вступ. У зв'язку з бурхливим розвитком технологій мультимедіа гостро постало питання захисту авторських прав та інтелектуальної власності даних, представлених у цифровому вигляді. Одним з найефективніших технічних засобів захисту мультимедійної інформації є цифрові водяні знаки (ЦВЗ), які застосовують переважно для захисту від копіювання і несанкціонованого використання.

Атаки проти стегосистем передачі повідомлень, що використовують ЦВЗ, стали більш поширеними і набули нових виглядів. А оскільки методи цифрових водяних знаків почали розроблятися цілком недавно, то тут є багато незрозумілих проблем, що потребують свого вирішення.

Матеріал і результати дослідження. Існують декілька видів атак на стегосистеми. Найбільш проста атака – суб'єктивна. Подібна атака може бути проведена лише проти зовсім незахищених стегосистем. Проте, вона, напевно, найпоширеніша на практиці, принаймні, на початковому етапі розкриття стегосистеми.

Первинний аналіз стегосистеми може містити в собі наступні заходи:

– Первинне сортування стегоповідомлень по зовнішніх ознаках.

– Виділення стегоповідомлень з відомим алгоритмом вбудовування.

– Визначення використаних стегоалгоритмів.

– Перевірка достатності обсягу матеріалу для стегоаналізу.

– Перевірка можливості проведення аналізу по окремих випадках.

– Аналітична розробка стегоматеріалів. Розробка методів розкриття стегосистеми.

– Виділення стегоповідомлень з відомими алгоритмами вбудовування, але невідомими ключами й т.д.

Розглянемо базову модель стегосистеми (рис.1), в якій в стегокодері використовується стеганографічна функція *f* вбудовування по секретному ключу *K* до приховуваного повідомлення *M* в контейнер *C*, а в стегодекодері стеганографічна функція *N* його витягання з використанням того ж ключа.

Із стего *S* по функції *F* витягується вбудоване повідомлення і при необхідності контейнер [1].

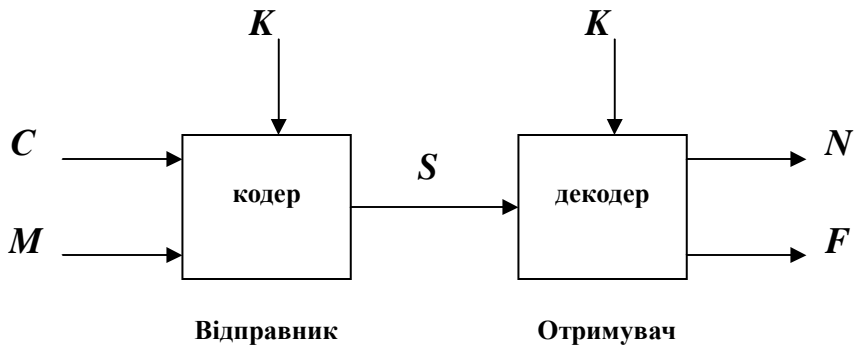

Рис. 1. Базова модель стегосистеми

В результаті спотворень при вбудовуванні, дії випадкових і навмисних перешкод передачі, а також погрішностей при витяганні цифрових даних, відновлене одержувачем повідомлення може відрізнятися від оригінала *M*. Аналогічно, отриманий контейнер відрізнятиметься від вихідного *C*. Контейнер обов'язково спотворюватиметься при вбудовуванні приховуваного повідомлення.

У ряді стегосистем необхідно відновлювати контейнер, оскільки він фізично є звичайними повідомленнями (зображення, мовні сигнали та інше) кореспондентів відкритого зв'язку, під прикриттям яких здійснюється прихований зв'язок.

За аналогією із криптоаналізом у стегоаналізі можна виділити наступні типи атак [2].

– Атака на основі відомого заповненого контейнера. У цьому випадку в порушника є одне або частина стегоповідомлення. В останньому випадку передбачається, що вбудовування схованої інформації здійснювалося відправником тим самим способом.

Завдання зловмисника може складатися у виявленні факту наявності стегоканала (основна), а також у його добуванні або визначення ключа. Знаючи ключ, порушник одержить можливість аналізу інших стегоповідомлень

– Атака на основі відомого убудованого повідомлення. Цей тип атаки більшою мірою характерний для систем захисту інтелектуальної власності, коли як водяний знак використається відомий логотип фірми.

Завданням аналізу є одержання ключа. Якщо відповідному схованому повідомленню заповнений контейнер невідомий, то завдання вкрай важко розв'язуване.

– Атака на основі обраного схованого повідомлення. У цьому випадку зловмисник має можливість запропонувати відправнику для передачі свої повідомлення й імітувати що виходять стегоповідомлення.

– Адаптивна атака на основі обраного схованого повідомлення. Ця атака є частним випадком попередньої. У цьому випадку зловмисник має можливість вибирати повідомлення для нав'язування відправнику адаптивно, залежно від результатів аналізу попередніх стегоповідомлень.

– Атака на основі обраного заповненого контейнера. Цей тип атаки більше характерний для систем ЦВЗ. Стегоаналітик має стегодетектор у вигляді «чорного ящика» і декілька стегоповідомлень. Аналізуючи детектовані сховані повідомлення, порушник намагається розкрити ключ.

Розглянуті види атак мають одну особливість: вони не змінюють стегоповідомлення, а також не спрямовані на протидію роботи декодера. У цьому полягає їхня позитивна сторона: дії зловмисника навряд чи здатні насторожити відправника й отримувача.

Стегосистеми з ЦВЗ повинні задовольняти суперечливим вимогам візуальної (аудіо) непомітності й дієздатності до основних операцій обробки повідомлень. Надалі без втрати спільності припустимо, що в якості контейнеру використовується зображення і система вбудовування повідомлень шляхом модифікації молодшого значущого біта (*LSB*) пікселів. Практично будь-який спосіб обробки зображень може призвести до руйнування значної частини вбудованого повідомлення.

Наприклад, розглянемо операцію обчислення ковзного середнього по двох сусіднім пікселям $(a+b)/2$, що є найпростішим прикладом низькочастотної фільтрації. Нехай значення пікселів a і b можуть бути парними або непарними з імовірністю $p=1/2$.

Тоді й значення молодшого значущого біта зміниться після усереднення в половині випадків.

До того ж ефекту може привести й зміна шкали квантування, скажемо, з 8 до 7 біт.

Аналогічний вплив робить і стиск зображень із втратами. Більше того, застосування методів очищення сигналів від шумів, що використають оцінювання й вирахування шуму, приведе до перекручування переважної більшості бітів схованого повідомлення.

Зрушення на декілька пікселів може призвести до не виявлення ЦВЗ у детекторі. Розглянемо це на прикладі стегоалгоритму [3].

У детекторі маємо $S_{WS} * W = (S_{0S} + W_S) * W = S_{0S} * W + W_S * W$,

де індексом S позначені зміщені версії відповідних сигналів.

Добуток $S_{0S} * W$, як і колись, близько до нуля.

Однак, якщо знаки $\pm$ у W вибиралися випадково й незалежно, то і $W_S * W$ буде близько до нуля, і стегоповідомлення не буде виявлено.

Всі атаки на стегосистеми з ЦВЗ можуть бути розділені на чотири групи [4]:

– атаки, спрямовані на видалення ЦВЗ;

– геометричні атаки, спрямовані на перекручування контейнера;

– криптографічні атаки;

– атаки проти використовуваного протоколу вбудовування й перевірки ЦВЗ.

Найпоширенішими є атаки, спрямовані на видалення ЦВЗ. До цієї групи ставляться такі атаки, як очищення сигналів-контейнерів від шумів, пере модуляція, стиск із втратами (квантування), усереднення й колізії. Ці атаки засновані на припущенні про те, що ЦВЗ є статистично описуваним шумом. Очищення від шуму полягає у фільтрації сигналу з використанням критеріїв максимальної правдоподібності або максимуму апостеріорної ймовірності.

Як фільтр, що реалізує критерій максимальної правдоподібності, може використатися медіанний (для ЦВЗ, що має розподіл Лапласа) або усереднюючий (для гауссовського розподілу) фільтр.

За критерієм максимуму апостеріорної ймовірності найкращим буде адаптивний фільтр Вінера (у випадку якщо як модель контейнера використається нестаціонарний гауссовський процес), а також граничні методи очищення від шуму (м'який і твердий пороги). Стиск із втратами й очищення сигналів від шумів значно зменшують пропускну здатність стегоканалу, особливо для гладких секторів зображення, коефіцієнти перетворення яких можуть бути «обнулені» без помітного зниження якості відновленого зображення.

Однак використання атак шляхом статистичного усереднення є особливо небезпечним, тому що в результаті знаходиться ЦВЗ і віднімається із зображення.

Такий вид атак можливий у випадку, коли атакуючий може одержати деякий узагальнений ЦВЗ, наприклад, деякий $W = f(S_0, W)$, незалежний сильно від вихідного зображення S_0.

Атакуючий може виявити ЦВЗ шляхом усереднення декількох зображень.

Наприклад, у нього є $S_0 + W, S_1 + W, ..., S_N + W$.

Тоді їхня сума $NW + \sum_j S_i$ буде досить близька до NM, якщо N велике, а зображення статистично незалежні.

Протидією проти подібної атаки може бути випадкове використання одного із двох ЦВЗ із ймовірностями p_1 й $p_2 = 1 - p_1$.

Тоді вище наведена атака дасть лише $p_1 W_1 + (1 - p_1) W_2$.

Однак, атака може бути поліпшена в тому випадку, якщо в атакуючі є якісь припущення про те, який ЦВЗ із двох убудований у дане зображення.

Тоді всі зображення можуть бути розподілені на два класи: 1 й 2.

Нехай $P_л$ – імовірність того, що зображення віднесене до невірного класу.

Тоді усереднення по великій кількості N_1 зображень класу 1 дає $x_1 = N_1 p_1 (1 - P_л) W_1 + N_1 (1 - p_1) P_л W_2$.

Аналогічне усереднення по N_2 зображень класу 2 дає $x_2 = N_2 p_1 P_л W_1 + N_2 (1 - p_1)(1 - P_л) W_2$.

Обчислення зваженої різниці дає

$$\frac{x_1}{N_1} - \frac{x_1}{N_2} = p_1 (1 - 2P_л) W_1 - (1 - p_1)(1 - 2P_л) W_2 .$$

Отже, для кожного $P_л \neq 1/2$, що атакує може оцінити суму й різницю $p_1 W_1$ й $(1 - p_1) W_2$, звідки він може одержати W_1 й W_2.

Різноманітність видів атак на стегосистеми с ЦВЗ підіймає актуальне питання про комплексний підхід до захисту таких систем з урахуванням сучасних напрямів розвитку інформаційних систем.

В рамкам даної роботи за мету ставиться дослідження можливості використання мультиагентної технології для захисту даних стегосистеми з ЦВЗ від атак.

Для реалізації поставленої мети необхідно виконати наступні завдання:

– провести порівняльний аналіз і вибір технологій мультиагентної системи (МАС) та дослідити моделі агентів;

– дослідити можливість використання технології МАС для захисту даних стегоситсеми з ЦВЗ від атак;

– здійснити вибір моделі бази знань за визначенням атак;

– створити базу знань (правив) відповідно до обраної моделі;

– розробити та відтестувати програмний додаток МАС до стегосистеми з ЦВЗ для захисту від атак.

Одним з головних завдань є вибір моделі бази знань. База знань може представляти експертні правила, отримані евристичним методом на основі накопиченого досвіду роботи з засобами захисту даних.

Модель бази знань може бути представлена у вигляді фреймів. Фрейм використовується для опису типової атаки на стегосистему з ЦВЗ. Слоти фрейма описують ознаки цієї атаки та посилання на процедури для відраження атаки [5].

Фрейми-екземпляри описують атаку певного виду. Процедурний слот описує дію або посилання на програмного агента, що виконує дію з відраження атаки. Таким чином, запропонований підхід використовує

продукційно-фреймову модель для створення мультиагентних застосувань по відображенню атак на стегосистему з ЦВЗ.

Евристичні правила бази знань за визначенням типу атаки можна описавати за допомогою правил-продукцій. В результаті виводів отриманих за допомогою бази знань локалізується джерело атаки і вибирається засіб захисту.

Висновки. Таким чином, еволюція атак на стегосистеми с ЦВЗ спонукає до пошуку нових рішень, щодо захисту таких систем. Тому використання мультиагентних технологій дає нові засоби вирішення цієї проблеми.

Реалізація завдання створення засобів МАС захисту стегосистеми з ЦВЗ зводиться до розробки програмного забезпечення, яке діє від імені системи, відповідно до політики безпеки стегосистеми і дозволяє відстежувати стан захищеності ЦВЗ, і своєчасно реагувати на виявлені атаки.

Література:

1.Понятие стеганографической стойкости [Електронний ресурс] // URL: http://crypts.ru/ponyatie-steganograficheskoj-stojkosti.html (дата: 10.04.2010)

2. Брюс Шнайдер. Прикладная криптография. 2-е изд. [Текст] – Москва: «Диа Софт», 2000. – 368 с.

3.Грибунин, В.Г., Оков, И.Н., Туринцев, И.В. «Цифровая стеганография» [Текст] – Москва: «СОЛОН-Пресс», 2002. – 272 с.

4.Классификация атак на стегосистемы ЦВЗ [Електронний ресурс] // URL: http://crypts.ru/klassifikaciya-atak-na-stegosistemy-cvz.html (дата: 13.04.2010)

5.Поспелов, Д.А. Искусственный интеллект. Кн. 2. Модели и методы: Справочник [Текст] – Москва: «Радио и связь», 1990. – 304 с.