DOI 10.15276/imms.v10.no 3-4.136
УДК 004.056.53

Informatics and Mathematical Methods in Simulation
Vol. 10 (2020), No. 3-4, pp. 136-141

РОЗРОБКА МЕТОДУ ПЕРЕДАЧІ СЕКРЕТНИХ ПОВІДОМЛЕНЬ НА БАЗІ ХЕШУВАННЯ ЦИФРОВИХ АУДІОФАЙЛІВ

В.В. Зоріло, О.Ю. Лебедєва, В.В. Конофольський

Одеський національний політехнічний університет,
пр. Шевченка, 1, Одеса, 65044, whiteswanhelena@gmail.com

Стеганографія та криптографія перебувають на піку розвитку в наші дні. Один з видів стеганографії – хеш-стеганографія. Це принципово новий напрямок в галузі передачі секретної інформації. Методи хеш-стеганографії засновані на тому, що замість вбудовування секретної інформації в контейнер, ми передаємо визначену послідовність цифрових сигналів та за їх хеш-кодами відновлюємо повідомлення. Для роботи за даними принципами зазвичай використовують цифрові зображення. Проте цифрові аудіо-файли завдяки різноманітним сервісам та месенджерам стають все популярнішим методом передачі інформації. Тому актуальною є розробка методів хеш-стеганографії, заснованих на використанні аудіо-сигналів. Метою даної роботи є підвищення ефективності передачі секретного повідомлення шляхом модифікації метода хеш-стеганографії. Даний метод не використовує вбудовування додаткової інформації в аудіофайл як контейнер. В даному методі запропоновано використовувати хеш-код (далі хеш), отриманий з аудіофайлів для побудови стеганоповідомлення. Але при передачі та конвертації аудіофайлів можливе порушення цілісності їх хеш-коду через можливі втрати якості і шуми. Це може стати перешкодою при встановленні відповідності отриманих файлів та їх хеш-кодів та при відновленні секретного повідомлення. Один з популярних сервісів ідентифікації аудіо-файлів – програмний додаток Shazam. Додаток Shazam дозволяє вірно ідентифікувати аудіо-треки навіть за наявності шумів, сторонніх звуків тощо. Розглянемо основні кроки даного алгоритму для використання їх при отриманні стійких до атак хеш-кодів аудіо-сигналів. Даний алгоритм дає змогу представити аудіо у вигляді набору особових точок, які далі методом MD5 перетворюються на хеш-коди. З даних хеш-кодів відновлюються секретні повідомлення після отримання визначеної послідовності аудіо-сигналів. Розроблений алгоритм не має аналогів для порівняння, проте є стійким до атак.

Ключові слова: хеш-стеганографія, аудіо-файл, хеш-код.

Вступ

Інформаційна безпека в наш час – особливо гостра та актуальна проблема. Інформатизація суспільства сягає таких масштабів, що важко знайти галузь людської діяльності, де б не використовували в тому чи іншому обсязі електронну інформацію. Разом з цим методи несанкціонованого доступу до цієї інформації, методи модифікації та знищення її також розвиваються дуже швидко. На заміну портативним комп'ютерам швидко ідуть сучасні смартфони, які часто є більш потужними за технічними параметрами, ніж деякі комп'ютери. Це дозволяє в будь-який момент мати доступ до певної інформації, або мати змогу «зламати» захист інформаційної системи та порушити інформаційну безпеку. Наслідки цих дій можуть бути як незначними, так і катастрофічними. Ситуація загострюється ще й через можливість працювати з великими масивами інформації та передавати цю інформацію в мережах. У сучасному світі передача великого обсягу інформації є нормою завдяки великій пропускній спроможності каналів зв'язку. Окрім цього, в певних галузях виникає необхідність передавати секретну або конфіденційну інформацію. Таку інформацію як правило не передають відкрито, для передачі використовують закриті канали зв'язку, або передають інформацію у зашифрованому вигляді. Для передачі секретних повідомлень

використовують криптографію, стеганографію та комбінацію тих і інших методів. Сучасні криптографічні методи є дуже надійними. Ефективність криптографічних алгоритмів визначається довжиною ключа. Сучасна техніка та інформаційні технології дозволяють зробити ключ настільки довгим, що підібрати ключ займатиме забагато часу. Сам злам втрачає свою актуальність через часові витрати на процес. Проте криптографія має і свої недоліки. При використанні криптографії в чистому вигляді зашифровані повідомлення привертають увагу. І хоч розшифрувати повідомлення, не маючи ключа, вкрай важко, або, навіть, неможливо, але сам канал зв'язку може бути розсекречено.

На зміну криптографічним методам прийшли стеганографічні. Стеганографія також передає зашифроване повідомлення, проте це повідомлення вбудовують у контейнер (цифрове відео, аудіо, цифрове зображення тощо). При передачі стеганоповідомлення самі контейнери не привертають уваги. Але і стеганографія має недоліки. При використанні стеганографічних методів завжди постає питання про пропускну спроможність сигналу-контейнеру та про стійкість метода до атак. Крім того, сьогодні часто використовують гібридні методи – у стеганографічний контейнер вбудовують повідомлення, зашифроване криптографічним алгоритмом. Таким чином значно підвищується ефективність засекречування. Але і недоліки у гібридних методів такі самі, як і в криптографічних та стеганографічних.

Однак сьогодні є альтарнатива, хеш-стеганографія. Хеш-стеганографія – передача повідомлення у вигляді послідовності цифрових файлів, хеш-коди яких в результаті аналізу формують повідомлення. Таким чином, не треба вбудовувати додаткову інформацію в контейнер. Послідовність контейнерів – це і є інформація.

У відкритому друці не знайдено робіт та посилань на використання аудіо-сигналів у хеш-стеганографії. Цифрові зображення – зручний та добре досліджений варіант. Оскільки аудіофайли є одним з основних форматів даних для передачі інформації в останні роки, питання приховування інформації в аудіофайл набуває все більшої актуальності. Існує безліч сервісів, що надають людям можливість прослуховувати і завантажувати аудіофайли різних форматів і якості, а також дедалі популярнішим стає спосіб спілкування голосовими повідомленнями. Усе це робить актуальним дослідження аудіо-файлів для передачі секретних повідомлень.

Мета даної роботи – підвищення ефективності передачі секретного повідомлення шляхом модифікації метода хеш-стеганографії.

Зараз хеш-стеганографія тільки починає свій розвиток. У відкритому друці знайдено роботи про вбудовування додаткової інформації в цифрові зображення: метод молодшого значущого біта, метод фазового кодування, методи розширення спектра [1], проте рідко згадуються методи хеш-стеганографії, засновані на використанні аудіо-сигналів.

Кодування методом молодшого значущого біта (LSB) – найпростіший спосіб вбудувати інформацію в цифровий аудіо-файл [2]. Замінюючи молодший біт кожної точки вибірки двійковим повідомленням кодування LSB дозволяє вбудувати великий обсяг даних. Однак в деяких реалізаціях кодування LSB два молодших біта вибірки замінюються двома бітами повідомлення. Це збільшує обсяг даних, які можуть бути закодовані, але також збільшує кількість результуючого шуму в файл-контейнер. Саме через шум методи LSB є нестійкими до всіх видів атак, а виявлення LSB-кодованого сигналу здійснюється за аномальними характеристикам розподілу значень діапазону молодших бітів [3]. Головною перевагою методу кодування LSB є низька обчислювальна складність алгоритму, в той час як його основним недоліком є те, що разом з тим, як збільшується кількість використовуваних найменших значущих бітів, збільшується й шанс виявлення застосування цього методу.

Фазове кодування вирішує недоліки шумових методів аудіостеганографії. Таке кодування спирається на той факт, що фазові компоненти звуку не так відчутні для

людського вуха, як шум. Замість того, щоб вводити спотворення, цей метод кодує біти повідомлення як фазові зсуви у фазовому спектрі цифрового сигналу, досягаючи нечутного кодування з точки зору відношення сигналу до сприйманого шуму. Цей метод має багато переваг перед LSB-кодуванням, найголовніша із яких є не виявляння для людського вуха. Як і всі методи, описані до цього часу, його слабкість все ще полягає в недостатній стійкості до змін в аудіо-файлах. Будь-яка окрема звукова операція або зміна даних призведе до спотворення інформації та запобігання її отриманню.

В аудіо-стеганографії метод розширеного спектру використовує для передачі секретних відомостей частотний спектр звукового сигналу. Проте, на відміну від методу LSB, метод розширеного спектру поширює секретну інформацію по спектру частот звукового файлу, використовуючи код, котрий не залежить від фактичного сигналу. В результаті кінцевий сигнал має високий рівень стійкості до атак, проте займає смугу пропускання, яка розміром більше, ніж необхідний розмір для передачі.

В роботі [4] описано метод хеш-стеганографії, заснований на використанні цифрових зображень. В основі даного методу лежить використання хеш-кодів зображення для формування послідовності зображень, кожне з яких відповідають певному символу секретного повідомлення (рис.1).

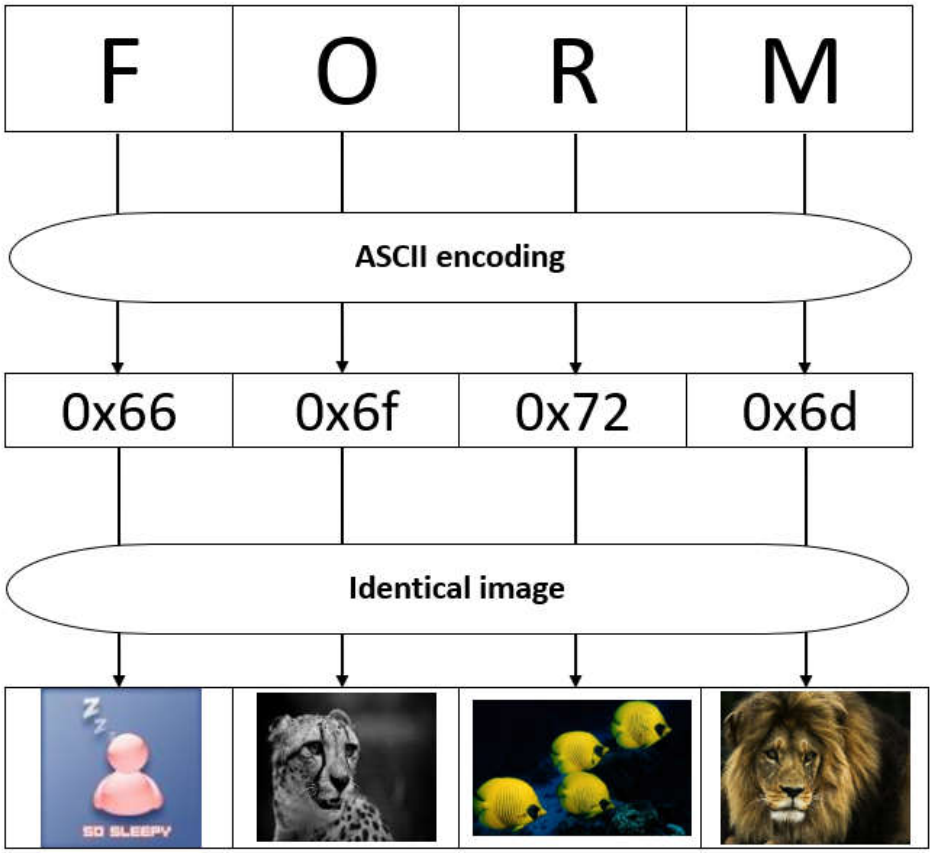

Рис. 1. Кодування повідомлення

Хеш-коди є дуже чутливими до будь-яких змін інформації. Ця їх властивість може стати перешкодою при передачі секретних повідомлень, адже при передачі інформації каналами зв'язку так чи інакше виникають шуми дискретизації тощо. Тому в роботі описано алгоритм модифікації використовуваних файлів перед обчисленням їх хеш-кодів таким чином, щоб вони були стійкими до атак. Описаний метод взято за основу в даній роботі. Його досліджено на цифрових зображеннях, даний метод має високу ефективність відновлення повідомлень навіть після застосування атак до

повідомлення. Проте проведемо дослідження на предмет використання замість цифрових зображень аудіо-файлів.

Основна частина

На відміну від перерахованих вище методів, даний метод не використовує вбудовування додаткової інформації в аудіофайл як контейнер у звичному розумінні. Повідомлення представляє собою послідовність аудіо файлів, хеш-коди яких після певних перетворень утворюють повідомлення, що передається.

В даному методі запропоновано використовувати хеш-код (далі хеш), отриманий з аудіофайлів для побудови стеганоповідомлення. Але при передачі та конвертації аудіофайлів можливе порушення цілісності їх хеш-коду через можливі втрати якості і шуми. Це може стати перешкодою при встановленні відповідності отриманих файлів та їх хеш-кодів та при відновленні секретного повідомлення. Один з популярних сервісів ідентифікації аудіо-файлів – програмний додаток Shazam. Додаток Shazam дозволяє вірно ідентифікувати аудіо-треки навіть за наявності шумів, сторонніх звуків тощо. Розглянемо основні кроки даного алгоритму для використання їх при отриманні стійких до атак хеш-кодів аудіо-сигналів.

1) Попередньо необхідно створити базу даних із зразків аудіо-файлів. Для цього буде використано мікрофон телефону, чи аудіо-гарнітури, підключеної до персонального комп’ютеру.

2) Зчитування та аналіз даних аудіо-зразків у байт-масиві. Для візуалізації даних байт-масив відобразимо у вигляді лінійного графіка (рис. 2).

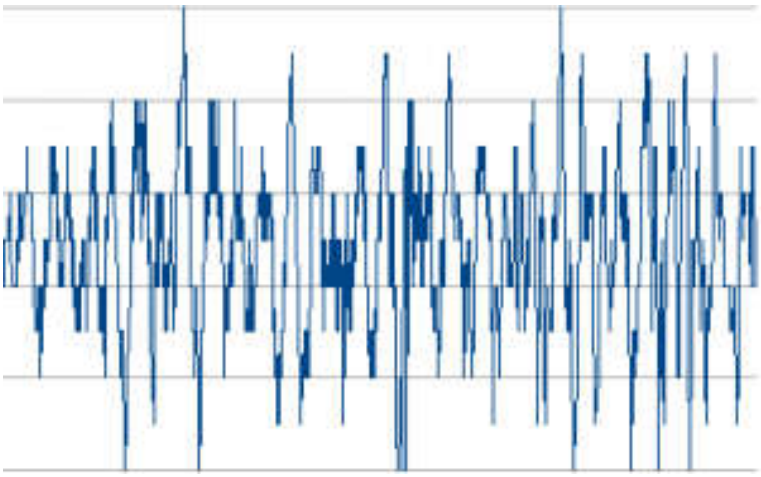

Рис. 2. Візуалізація отриманих даних

Ці дані зазвичай називають часовою областю. Представлення звукового файлу в такому вигляді дозволяє проводити його аналіз у відношенні до часу [4]. Але отримання цих даних є лише початком, бо для алгоритму Shazam є необхідним отримання спектрального аналізу замість прямих даних часової області.

3) Спектральний аналіз. На даному етапі до файлу застосовують дискретне перетворення Фур’є, що перетворить дані с часової області в частотну. Коли дані перетворюються в частотну область, втрачається кожний біт інформації про час. Таким чином, отримується величина всіх частот, але без інформації про те, коли вони з’являються. Для вирішення цієї проблеми дані розподіляються на сегменти та перетворюються у цих сегментах по одному. Завдяки цьому можна визначити дані всіх частот на утвореному сегменті.

4) Визначення ключових точок в аудіофайлі та збереження їх у вигляді хеш-коду. Для кожної лінії спектрального аналізу обираються точки з найбільшою величиною із певного діапазону (рис. 3).

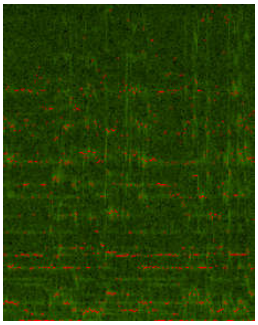

Рис. 3. Отримання ключових точок аудіо-файлу

5) Отримання «зовнішніх» аудіофайлів та порівняння їх з існуючими у базі даних за допомогою отриманих ключових точок.

Розглянемо можливість застосування даного алгоритму з метою передати секретне повідомлення через послідовність аудіо-файлів. Створимо базу даних з аудіофайлів і отриманих з них хеш-кодів. Для бази даних необхідна велика кількість унікальних аудіофайлів. Хеш-код аудіофайлів отримуємо наступним чином.

1. За допомогою алгоритму Shazam отримуємо ключові точки аудіо-файлу, значення яких переведемо у тип цілих чисел. Далі для послідовності отриманих чисел обчислимо хеш-код за допомогою алгоритму MD5. Одному символу повідомлення поставимо у відповідність чотири зразки аудіо-сигналів таким чином, щоб ASCII-код символу повідомлення співпадав з двома першими символами хеш-кодів аудіофайлів (рис. 4).

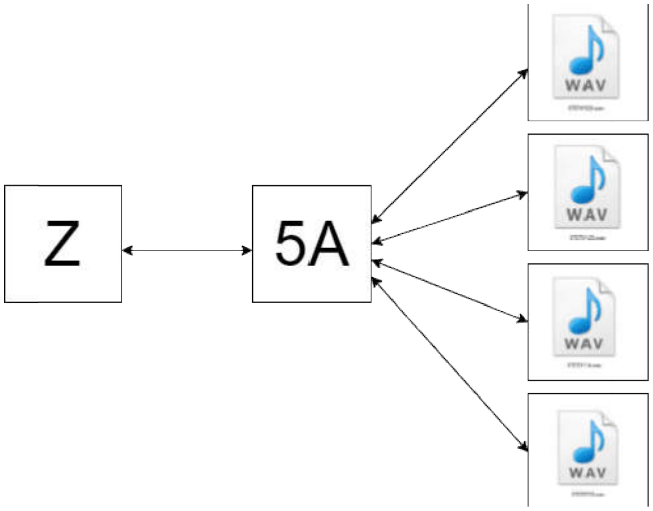

Рис. 4. Схема відповідності «символ-аудіо»

2. Перетворюємо наше повідомлення у ASCII-послідовність та поділяємо її на рівні частини по два символи.

3. Кожній парі символів ASCII-послідовності ставимо у відповідність аудіо-файл. Отримаємо послідовність аудіо-файлів, що є контейнером стеганоповідомлення та готова до передачі.

4. Після отримання послідовності аудіо-файлів одержувач записує у ряд перші два символи з хеш-кодів файлів, в результаті чого отримує ASCII-послідовність, котру необхідно перевести в текст для отримання оригінального повідомлення.

Висновки

В роботі представлено метод хеш-стеганографії, заснований на використанні хеш-кодів цифрових аудіо-сигналів. Алгоритм даного методу є стійким до різноманітних видів атак та на даний час не має аналогів для порівняння. Подальші зусилля авторів спрямовано на дослідження можливостей даного методу та підвищення його ефективності.

Список літератури

1. Иваненко В.Г., Пивошенко Я.И. Способ защиты авторского права на аудио сигналы основанный на пакетной вейвлет-декомпозиции. *Безопасность информационных технологий.* 2013. № 1. С. 72–74.
2. Roy S., Manasmita M. A novel approach to format-based test stegonography, *International conference on communication computing and security*, ICCCS 2011, Proceedings by ACM, Odisha,India. P.511-516
3. Грибунин В.Г., Оков И.Н., Туринцев И.В. Цифровая стеганография. М.: Солон-Пресс, 2002. 272 с.
4. Бохонько М.В., Зоріло В.В. Модифікація методу хеш-стеганографії. *VII Міжнародна науково-практична інтернет-конференція «Сучасний рух науки»*, 6-7 червня 2019 р. Дніпро, 2019. С.605-608.

DEVELOPMENT OF A METHOD FOR TRANSMITTING SECRET MESSAGES BASED ON HASHING OF DIGITAL AUDIO FILES

V.V. Zorilo, O.Yu. Lebedeva, V.V. Konofolskyi

Odessa National Polytechnic University,
1, Shevchenko Ave, Odessa, 65044, Ukraine, whiteswanhelena@gmail.com

Steganography and cryptography are at their peak these days. One of the types of steganography is hash steganography. This is a fundamentally new direction in the transmission of secret information. Hash steganography methods are based on the fact that instead of embedding secret information into a container, we transmit a specific sequence of digital signals and recover messages from their hash codes. Digital images are commonly used for these principles to work. However, digital audio files, thanks to the many services and messengers, are becoming an increasingly popular method of transferring information. Therefore, the development of hash steganography methods based on the use of audio signals is urgent. The aim of this work is to improve the efficiency of transmission of a secret message by modifying the hash steganography method. In this paper, the Shazam`s algorithm is used to obtain hash codes of a digital audio signals that are resistant to attacks and file transformations. this method does not use embedding additional information in the audio file as a container. In this method, it is proposed to use a hash code (hereinafter the hash) obtained from audio files to build a stegan message. But when transferring and converting audio files, the integrity of their hash code may be compromised due to possible loss of quality and noise. This can interfere with matching files and hash codes and restoring the secret message. One of the popular audio file identification services is the Shazam software application. Shazam allows you to correctly identify audio tracks, even in the presence of noise, extraneous sounds and more. Consider the main steps of this algorithm for their use in obtaining attack-resistant hash codes of audio signals. This algorithm allows representing audio as a set of special points, which are then converted into hash codes using the MD5 method. Secret messages are recovered from these hash codes after receiving a specific sequence of audio signals. The developed algorithm has no analogues for comparison, but it is resistant to attacks.

Keywords: hash-steganography, audio-file, hash-code.