УДК 004.056: 621.397

В.Ю. КОРОЛЬОВ, В.В. ПОЛІНОВСЬКИЙ, В.А. ГЕРАСИМЕНКО, М.Л. ГОРІНШТЕЙН
Інститут кібернетики ім. В.М. Глушкова НАНУ, м. Київ

ПЛАНУВАННЯ ДОСЛІДЖЕНЬ МЕТОДІВ СТЕГАНОГРАФІЇ ТА СТЕГОАНАЛІЗУ

Запропоновано концепцію комп'ютеризованих стегоаналітичних досліджень методів стеганографії та сформульовано вимоги до подібних систем. Показано, що розроблене програмне забезпечення дозволяє отримати комплексну оцінку можливостей методів приховування даних у зображеннях. Результати роботи програмного комплексу показані на прикладі RS-стегоаналізу для НЗБ-алгоритмів приховування даних.

The concept of new information technology research methods for stegoanalylsis and formulated requirements for such systems are given. Shown that the developed software allows to obtain a comprehensive assessment of the possibilities of methods for hiding data in images. Results of the software system shown on the RS-stegoanalysis for LSB-algorithms of hidden data detection.

Keywords: steganography, RS-stegoanalysis, information technologies.

Вступ. Сучасні комп'ютерні технології обробки даних дозволили широко використовувати криптографічні методи захисту інформації. Однак, для ряду прикладних завдань інформаційної безпеки криптографічних методів стає недостатньо, оскільки вони не дозволяють приховати сам факт наявності або передачі інформації з обмеженим доступом. В таких випадках актуальними стають стеганографічні методи.

Комп'ютерна стеганографія (КС) активно розвивається вже понад 20 років. Як відомо, на відміну від криптографічного захисту інформації стеганографічні програмні засоби [1, 2] намагаються, в першу чергу, приховати сам факт передачі даних. Найчастіше в якості носія для приховування додаткової таємної інформації слугують мультимедійні файли (контейнери), КС використовує у своїх методах їх психовізуальну надлишковість. Водночас, сучасна КС використовує методи криптографії для шифрування інформації перед її вбудовуванням в контейнер, що з точки зору статистики еквівалентно внесенню у контейнер стохастичного збурення. Таке вбудовування може застосовуватись як метод захисту авторських прав, а саме впровадженням цифрових водяних знаків (ЦВЗ), так і як метод для приховування інформації з метою запобігання її викрадення або замаскованої передачі.

Спрощення використання методів приховування інформації та можливість передачі інформації по відкритим цифровими каналам передачі даних зробили доступними стеганографічне програмне забезпечення пересічному користувачу персонального комп'ютера з доступом до мережі Інтернет. Сьогодні на ринку програмного забезпечення існує багато пропозицій стеганографічних програмних додатків, у тому числі й на безоплатній основі.

Зрозуміло, що засоби стеганографії можуть використовуватись як законослухняними громадянами, так і кримінальними або шпигунськими структурами, тому активно розвиваються відповідні методи протидії – стегоаналіз, які покликані виявити приховану у контейнері інформацію, або виявлення факту прихованої передачі даних.

Комп'ютерний стегоаналіз. Стегоналіз є пасивною атакою на стеганографічні системи, тобто такою, що не змінює зміст повідомлення. Сьогодні комп'ютерний стегоаналіз (СА) виділяється як самостійний науковий напрямок, метою якого є виявлення в носії (контейнері) прихованих даних і оцінка об'єму цих даних. СА широко використовує апарат математичної статистики, лінійної алгебри, комбінаторики, теорії планування експерименту, статистичного аналізу, цифрової обробки й розпізнавання сигналів і зображень, а також інші розділи математики. Чисельні експерименти по впровадженню й виявленню прихованих даних є основним способом одержання достовірних відомостей про якість роботи алгоритмів стеганографії та стеганализу. Дослідження стійкості методів приховування даних до СА дозволяє перевірити надійність стеганографічних алгоритмів, а також зробити вагомий внесок в інформаційну безпеку держави.

Аналіз останніх досліджень і публікацій

За останні 15 років створено багато методів приховування інформації у різних типах і форматах файлів [1-15], а також методів виявлення вбудованих даних. Найпопулярнішим на сьогодні методом стеганографічного приховування є метод заміни найменш значущих біт (НЗБ-стеганографія). Ідея методу полягає в заміні від одного до чотирьох молодших бітів в байтах кольорів пікселів початкового зображення бітами приховуваного повідомлення [10]. Можливість такої заміни, як вже зазначалось, обумовлена наявністю в зображеннях структурної надлишковості. Метод застосовується до растрових зображень, представлених у форматах без втрат. Одним з таких форматів виступає BMP. Позитивною стороною BMP є висока якість зображення, а також простота формату, що робить його популярним для застосування у якості контейнеру.

Відповідно, найбільше методів стегоаналізу розроблено для виявлення саме НЗБ-стеганографії [1-15]. Крім того, такі методи найбільш прості для програмування. Тому більшість комерційних і вільних програм приховування даних мають у своєму складі додатки НЗБ-стеганографії. Одним з найбільш точних сучасних методів виявлення прихованих даних у зображеннях, збережених у форматах без втрат, є RS-стегоаналіз [3-15].

Проте абсолютна більшість сучасних цифрових фотографій зберігається у форматі JPEG, оскільки у ньому найкраще реалізовано стиск зображень при мінімумі втрат візуальної якості. Висока продуктивність JPEG-алгоритмів ґрунтується на швидких перетвореннях у яких обмежується інтенсивність високочастотних складових зображень. Але як стверджується у роботі [3], безпосереднє застосування методів НЗБ-стеганографії до зображень у форматі JPEG достатньо просто може бути виявлено, оскільки суттєво спотворює співвідношення чисел зображення у такому форматі і тому використання файлів збереження цифрових фотографій з розширеннями крім JPEG для передачі або демонстрації аматорських фотографій через Інтернет одразу стає підозрілим з позиції стегоаналітика.

Формат RAW (цифровий негатив), у якому зберігають зображення професійні і напівпрофесійні дзеркальні фотокамери за побудовою не призначений для перегляду кольорових фотографій, оскільки відображає неструктуровану інтенсивність оптичної густини світлового потоку з ПЗЗ-матриці фотоапарата, тобто безпосередній перегляд зображень можливий лише у монохромному виді (відтінках сірого, чорно-білому).

При створенні методів стегоаналізу розробники виходять з того, що користувачі будуть фотографувати об’єкти або сцени фотокамерами середнього чи високого класу, або братимуть цифрові фотографії з тематичних сайтів у мережі Інтернет. Проте зображення може бути оброблене власником фотографії (обробка зображення для демонстрації в Інтернет), або самим користувачем для того щоб унеможливити для стегоаналітика отримання точного оригіналу.

Зазначимо, що тенденцією останніх двох років сучасного стегоаналізу [12-14] є використання методів інтелектуального аналізу даних (ІАД – Data Mining) у стегоаналізі. Для чого з декількох методів стегоаналізу і математичної статистики виділяють групи ознак для виявлення прихованих даних, які потім передають у системи ІАД.

Проблема дослідження статистичних властивостей контейнерів

Раціональна комплектація вибірок цифрових фотографій дозволяє розширити область дослідження, точніше розкрити закономірності функціонування, скоротити потреби в обчислювальних ресурсах на проведення досліджень. Підготовка чисельних статистичних експериментів для стегоаналіза (СА) полягає у формуванні із сукупності доступних масивів фотографій, класифікованих за певними ознаками вибірок, за заздалегідь складеним планом. Детальне планування досліджень з метою оптимізації параметрів методів стеганографії та стегоаналізу дозволяє одержати оцінки ступеня правдоподібності показників, заявлених розробниками методів СА, виявити закономірності впливу методів вбудовування даних на глобальні й локальні статистичні закономірності, проаналізувати вплив сукупності параметрів фотографії й умов експозиції.

Збір статистичних даних методом пасивного спостереження за результатами чисельних експериментів з неструктурованою сукупністю фотографій вимагає значних витрат обчислювальних ресурсів і високопродуктивних систем зберігання даних. Перераховані причини змушують приділяти серйозну увагу раціональній організації експериментального вивчення таких об’єктів. Ціль представленої роботи полягає в розробці раціональних і доцільних планів проведення чисельних експериментів з метою скорочення обсягу проведених досліджень заданої точності й вірогідності одержання достовірних результатів, добування з отриманих даних максимуму корисних відомостей та розробки протоколів верифікації методів СА.

Постановка завдання

Як було зазначено вище, планування статистичних досліджень для RS-стегоаналізу і подібних методів є актуальною науковою проблемою. Одним із завдань якої є дослідження стійкості методів приховування даних до стегоаналізу. У роботі викладено результати понад шести років науково-прикладних досліджень авторів роботи в області НЗБ-стеганографії, які можуть бути застосовані як для приховування в контейнерах у форматі JPEG, так і для методів, які є подальшим розвитком RS – WS-стегоаналіз [3].

Основний розділ

Етапи стегоаналітичного дослідження. Досвід авторів у проведенні стегоаналітичних досліджень показує, що розробники методів стегоаналізу (СА) і автори наукових робіт, які намагаються знайти в цих методах слабкі сторони і виявити їх обмеження, не достатньо уваги приділяють формуванню колекції файлів, з якими виконуються дослідження. Як правило, з мережі Інтернет для статистичних досліджень беруть набори фотографій (кількістю 150-450 файлів) які, на думку авторів методів, розміщені там без оброблення цифровими фільтрами і без вбудовування знаків захисту авторського права (ЦВЗ).

У нашому циклі робіт було показано [7-10], що такий підхід приводить до перебільшення можливостей методу стегоаналізу. Крім того, на результати статистичних досліджень впливає не тільки ступінь зашумлення зображення, але і геометричний розмір та розрізнення фотографії. Так для зображень більшого формату без вбудованих стеганографічних даних характерно менше значення хибно позитивно виявлених стегобіт (ХПВС) [9] – величини природного шуму та артефактів реєстрації фотографій, які невірно трактуються методом RS-стегоаналізу як приховані дані. З іншого боку, метод СА при моделюванні роботи НЗБ-стеганографічного алгоритму видаватиме величину позитивно виявлених стегобіт (ПВС) з похибкою в більшу або в меншу сторону в силу стохастичної природи формування зображення і випадкових збігів та неспівпадінь значень стегобіт з бітами цифрової фотографії.

Проблема планування СА

Велика кількість методик стеганалізу не виключає необхідності детального визначення параметрів

процесів експозиції фотографій на цифрових пристроях, аналізу умов зйомки, вивчення факторів, що впливають на об'єкт у процесі реєстрації. Ці відомості потрібні на етапі складання плану експерименту, при аналізі й інтерпретації результатів. Матеріал викладений на рівні, призначеному в основному для науково-прикладних досліджень і не ставить метою математично строгий доказ розглянутого підходу. Для задачі СА цілком природно застосувати критерій мінімуму числа експериментів, тобто серед усіх планів бажано обирати такий, який вимагає мінімального числа дослідів при дотриманні вимог до якості статистики або її параметрів.

У відповідності до теорії планування експерименту [16] на початковому етапі досліджень, коли немає відомостей про вплив тих чи інших параметрів на мету досліджень, необхідно виконати відсіювальні експерименти для виявлення параметрів, які не суттєво впливають на статистичні дані або породжують аномальні відгуки, що рідко зустрічаються на практиці. Відсіювання несуттєвих факторів дозволяє знизити трудомісткість задач СА.

Кожна нова інформаційна технологія при широкому використанні суспільством проходить етап стандартизації, метою якого є верифікація зразків, наданих виробником. За аналогією до криптографічного ПЗ, стеганографічні програми також мають проходити перевірку, для чого слід розробити відповідні протоколи верифікації та плани випробувань.

На думку авторів, **спрощений початковий план стегоаналітичних досліджень** повинен складатись з наступних етапів.

1) Комплектування наборів цифровими зображеннями для стегоаналітичних досліджень, які відповідають наступним вимогам:

А) Зображення у вибірках (директоріях) мають бути з одного джерела (пристрою реєстрації, одного походження: фотоапарата, сканера і т.п.);

Б) Зображення мають бути отримані в приблизно однакових умовах і подібного змісту (тематики або фотографованої сцени);

В) Зображення з приблизно рівними геометричними розмірами у пікселях та однаковою роздільною здатністю (dpi – dot per inch, кількість точок на дюйм);

Г) Налаштування пристрою реєстрації (фотоапарату) при створенні колекцій не повинні суттєво відрізнятись: витримка, величина діафрагми, чутливість до світла тощо.

Завдяки виконанню цих вимог спрощується виявлення закономірностей і слабких місць стегоаналізу, що є основою для побудови методів обходу, відповідно трактовка результатів стає більш прозорою і однозначною.

2) Цільове формування об'ємів вибірок зображень. На різних етапах дослідження методу стегоаналізу потрібні вибірки різного об'єму (кількістю файлів), щоб скоротити час отримання результату. З нашого досвіду та з досвіду інших дослідників можна рекомендувати наступні значення:

А) Вибірки для перевірки якості роботи програми (10-20 файлів);

Б) Вибірки для первинних досліджень (десятки файлів);

В) Вибірки для перевірки гіпотез та методів обходу стегоаналізу (200-300 файлів);

Г) Вибірки великого об'єму для остаточного підтвердження статистичної достовірності висунутих гіпотез (тисячі файлів – десятки тисяч файлів);

Д) Вибірки з різними (певними) законами розподілу інформації (для пошуку ідеального розподілу), кількість зображень декілька тисяч.

Формалізація напрямків статистичних досліджень RS-стегоаналізу

Дослідження RS-СА можна згрупувати у три основні напрямки (рис. 1). Розглянемо їх більш детально:

1) Кількісні дослідження – збір статистики для зображень з вбудованими даними і для оригіналів (без прихованих даних). На первинному етапі пошуку гіпотетичних слабких місць методу рекомендується використовувати неструктуровані набори фотографій та зображень (в тому числі штучного походження, наприклад, синтезованих) з різноманітними параметрами і характеристиками. За отриманими даними (особливостями зображень – аномалій) можна зробити правдоподібні припущення, щодо помилкових виявлень прихованих даних та інших слабких місць методу стегоаналізу.

2) Визначення впливу цифрової фільтрації на RS-СА результати. Аналіз існуючих операцій над зображеннями при підготовці до друку або та типових операцій заглушення шуму і покращення візуальної якості: посилення яскравості і підвищення контрасту тощо. Технічно реалізується за допомогою модульної архітектури програмного комплексу, що описана нижче.

3) Дослідження якісно-кількісних співвідношень зображень – нерівностей (статистики високого порядку, похідні статистичні характеристики)

4) Суміжні дослідження: зв'язок RS-СА з середньоквадратичним відхиленням (СКВ) шуму при додаванні стегобіт, нерівності для медіан, середніх значень, цифрових водяних знаків (ЦВЗ), співвідношення між байтами кольорових каналів [9].

5) Тестування методів обходу: попередня фільтрація, підбір зображень додавання в які стегобіт не збільшує суттєво відсоток ПВС за RS-СА.

6) Створення тегів для зображень. Мета шостого етапу полягає у семантичному описі змісту кожного зображення з досліджуваної колекції для передачі його в систему інтелектуального аналізу даних (ІАД – Data Mining).

7) Застосування методів ІАД для виявлення закономірностей не помічених дослідниками СА. Одним з найбільш відомих програмних комплексів ІАД є програмний комплекс Data Mining фірми Oracle.

Архітектура програмного комплексу аналізу зображень

Для досліджень характеристик масивів зображень за різними методами було розроблено універсальний комплекс аналізу зображень з модульною архітектурою, який дозволяє додавати нові формати файлів зображень та алгоритмів їх аналізу, а також виконувати обробку без зміни самого комплексу. Головним завданням, яке ставилось при розробці комплексу стегоаналізу, була потреба виконання досліджень не лише оригінальних масивів зображень, а й певним чином модифікованих версій у процесі обробки зображень, наприклад – результати фільтрації за обраними оператором-аналітиком алгоритмами. Крім того, необхідно виконувати аналіз зображень різними алгоритмами, тобто в загальному випадку над зображеннями необхідно виконати задану послідовність операцій фільтрації та аналізу. Кожна операція повинна мати можливість налаштування, тобто завдання параметрів фільтрації та аналізу.

Рис. 1. Структурна схема напрямків досліджень методу СА

Отримані масиви статистик зберігаються в базі даних, структура якої дозволяє зберігати та отримувати статистику оброблену іншими модулями аналізу даних. Також комплекс має досить широкі можливості вибірки та аналізу накопичених даних. За переліченими характеристиками комплекс стегоаналітичних досліджень суттєво відрізняється від наявних рішень, більшість з яких є вузькоспеціалізованими і виконують аналіз зображень лише одного типу та одним алгоритмом, результати обробки також представляються в своєму форматі.

Для розв`язання цієї задачі була реалізована модульна архітектура комплексу, яка передбачає реалізацію фільтрів та аналізаторів в окремих модулях-розширеннях комплексу. Це дозволяє додавати такі модулі без зміни основного комплексу. Фільтр та аналізатор в цій архітектурі визначаються як програмні інтерфейси (набір методів з визначеними сигнатурами), що використовуються основним комплексом при обробці файлів. Модулі-розширення представляють собою звичайні dll-бібліотеки, що містять один чи декілька класів, які реалізують ці інтерфейси.

Серед прототипів комплексу можна назвати комплекси MGEBO [12] та Digital Invisible Ink Toolkit [13], а серед аналогів – додатки розроблені лабораторією проф. Д. Фридрич [14], Virtual Steganographic Laboratory for Digital Images [15].

На рис. 2 наведена діаграма класів поточної версії комплексу, яка включає інтерфейси фільтру та аналізатору, а також декілька класів, що реалізують ці інтерфейси. Кожен з цих класів реалізований в окремому проекті dll-бібліотеки.

Крім інтерфейсів на діаграмі наведені 2 класи фільтрів та 4 класи аналізаторів:

- ImageFilters – реалізує набір стандартних графічних фільтрів: GaussianBlur, Defocus, Highlight, Sharpen, BigEdge, Emboss, EmbossColor, EdgeDetect, Negative, RemoveChannel, Punch;
- SteganosFilter – реалізує фільтри приховування даних, які базуються на алгоритмі НЗБ;
- MathStatAnalyzers – виконує аналіз зображень методами математичної статистики, повертаючи середнє значення, середньоквадратичне відхилення та медіану для декількох характеристик з різних кольорових просторів;
- RS-Analyzer – виконує RS-аналіз зображення та повертає набір відповідних коефіцієнтів;
- ColorComparisonAnalyzer – виконує порівняння кольорових складових пікселів зображення (R, G, B) та повертає статистику по співвідношенням між ними;
- NoiseAnalyzer – повертає дві шумові характеристики для різних кольорових просторів.

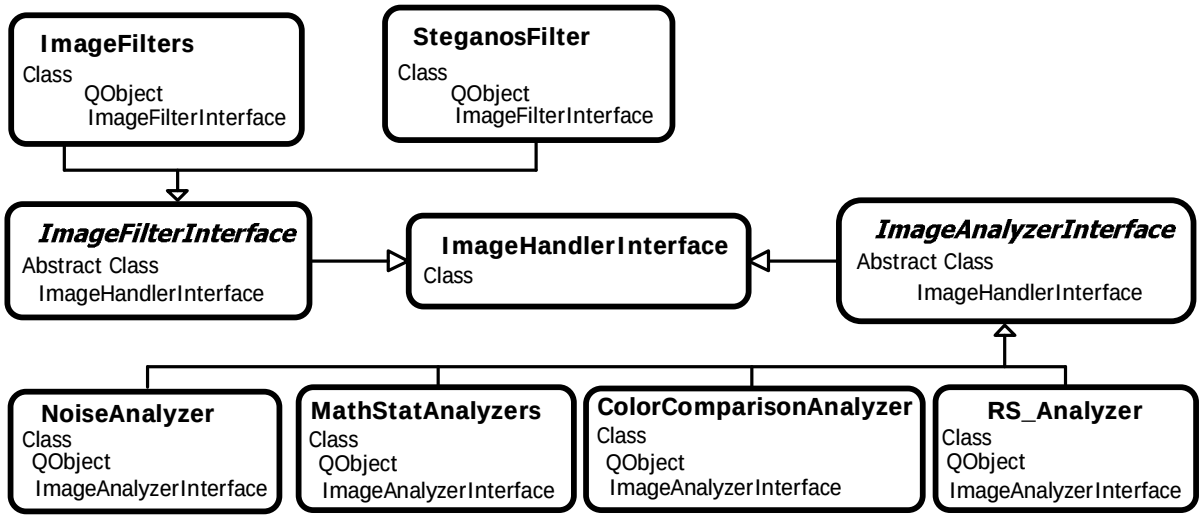


Рис. 2. Діаграма класів фільтрів та аналізаторів

Кількість таких класів фільтрації та аналізу буде розширюватись для підтримки нових методів та алгоритмів. Робота з комплексом розбивається на декілька етапів:

1) Вибір файлу бази даних, в якому буде зберігатись статистика.

2) Додавання папок із зображеннями, які необхідно обробити. При цьому підтримується рекурсивна обробка піддиректорій та завдання списку масок файлів для обробки.

3) Завдання послідовності екземплярів фільтрів та аналізаторів, якими будуть оброблятись та досліджуватись фотографії. Фільтр – це модуль, який змінює зображення, наприклад, виконує його фільтрацію чи приховування даних певним алгоритмом. Аналізатор повертає певний набір статистики. Для модулів обох типів можуть задаватися набір специфічних для них параметрів (наприклад, коефіцієнти роботи алгоритмів фільтрації та аналізу) – таким чином створюються екземпляри фільтрів та аналізаторів, які і додаються в послідовність.

4) Запускається на виконання завдання обробки файлів по даним попередніх етапів. Реалізація обробки виконана по схемі робочих потоків, які незалежно оброблюють файли зображень, завдяки чому підвищено ефективність паралельної обробки на SMP-системах. Підтримка розподілених систем планується в наступних версіях комплексу.

5) Після закінчення обробки накопичена статистика доступна для вибірок та експорту у вигляді звітів трьох типів (дивись рис. 3).

Статистика по зображенням дозволяє отримати результати аналізу зображень по кожному файлу окремо. При цьому можна вибрати необхідний набір даних, які будуть виводитись для зображень, а також задати умову фільтрації по ним, наприклад: ([Ширина] > 1000) and ([Висота] > 1000) and ([І'мя] like 'nature %'). На другій вкладці можна вибрати набір даних статистики, яку необхідно вивести. Набір доступних даних статистики відображається у вигляді дерева, яке містить послідовності екземплярів фільтрів та аналізаторів з вкладеними списками статистики, яку можна включити в звіт (рис. 4).

По даним статистики також можна виконувати фільтрацію, задаючи потрібним колонкам символічні імена та використовуючи їх у виразі фільтру, наприклад: ([a1] > 90) and ([a2] > 40). Після завдання параметрів звіту можна переглянути результат на третій вкладці. Його можна експортувати в csv-файл для подальшої обробки в табличному процесорі.

Другий варіант статистики – сукупний, дозволяє виводити агреговану вибраною функцією (мінімум, максимум, сума, кількість та середнє значення) статистику по заданим даним, згруповану по вибраним колонкам. Наприклад, можна отримати сумарну статистику окремо по всім папкам, в яких знаходяться оброблені зображення.

Третій варіант статистики дозволяє отримати частоти появи трійок, що відображають співвідношення трьох вибраних величин. Статистика по зображенням дозволяє отримати результати аналізу зображень по кожному файлу окремо. При цьому можна

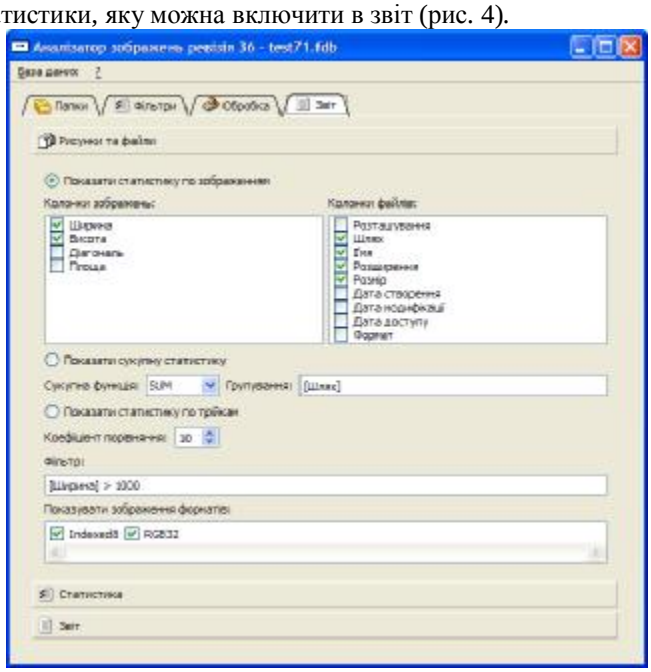


Рис. 3. Параметри отримання звіту зі статистикою

вибрати необхідний набір даних, які будуть виводитись для зображень, а також задати умову фільтрації по ним, наприклад: ([Ширина] > 1000) and ([Висота] > 1000) and ([Ім'я] like 'nature %'). На другій вкладці можна вибрати набір даних статистики, яку необхідно вивести. Набір доступних даних статистики відображається у вигляді дерева, яке містить послідовності екземплярів фільтрів та аналізаторів з вкладеними списками статистики, яку можна включити в звіт. За даними статистики також можна виконувати фільтрацію, задаючи потрібним колонкам символічні імена та використовуючи їх у виразі фільтру, наприклад: ([a1] > 90) and ([a2] > 40). Після завдання параметрів звіту можна переглянути результат на третій вкладці. Його можна експортувати в csv-файл для подальшої обробки в табличному процесорі типу Майкрософт Ексель.

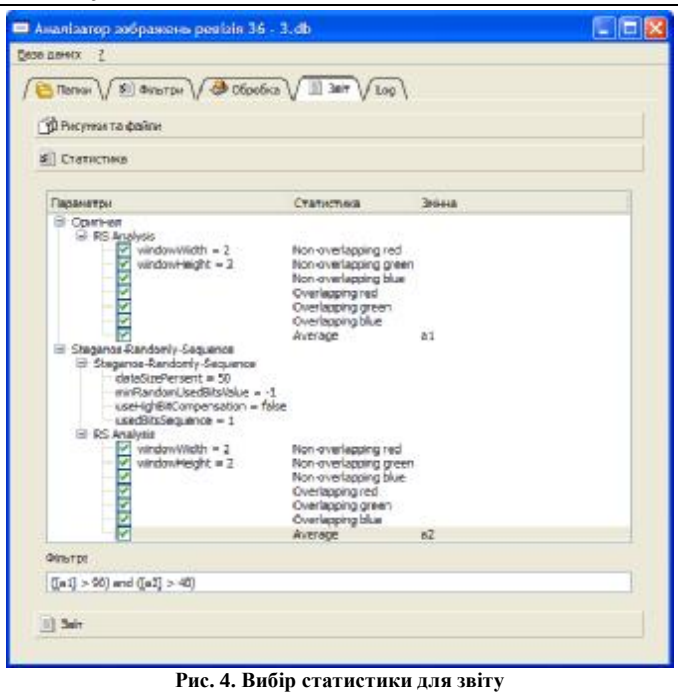


Рис. 4. Вибір статистики для звіту

Приклади результатів досліджень

Відомо, що зображення з мереж загального доступу можуть мати цифрові водяні знаки (ЦВЗ), над ними можуть бути виконані операції покращення візуальної якості або спеціальні дизайнерські перетворення типу підготовки до друку. Метою дослідження є аналіз впливу ЦВЗ на результати RS-CA.

Дослідження впливу вбудовування ЦВЗ на результати RS-стегоаналізу

Цифровий водяний знак (ЦВЗ) – спеціальна мітка, непомітно впроваджувана в зображення або інший сигнал з метою тим або іншим способом контролювати його використання. За допомогою плагину Digimarc ПЗ Adobe Photoshop та програми Image Watermarks в досліджувані зображення впроваджувалися текстові та графічні ЦВЗ. Результати впливу ЦВЗ на зображення представлені в табл. 1.

При дослідженні використовувались декілька варіантів текстових та графічних ЦВЗ. В таблиці представлені середні значення отриманих результатів.

Таблиця 1

Вплив ЦВЗ на результати RS-CA

	Кількість зображень, %				
	0-2	2 – max			
		2 – max	2 – 5	5 – 10	10 – max
Оригінал	**80,7**	**19,3**	**13**	**4,6**	**1,7**
текст	73,5	26,6	18,8	6,1	1,7
зображення	78,5	21,5	14,5	5,5	1,5

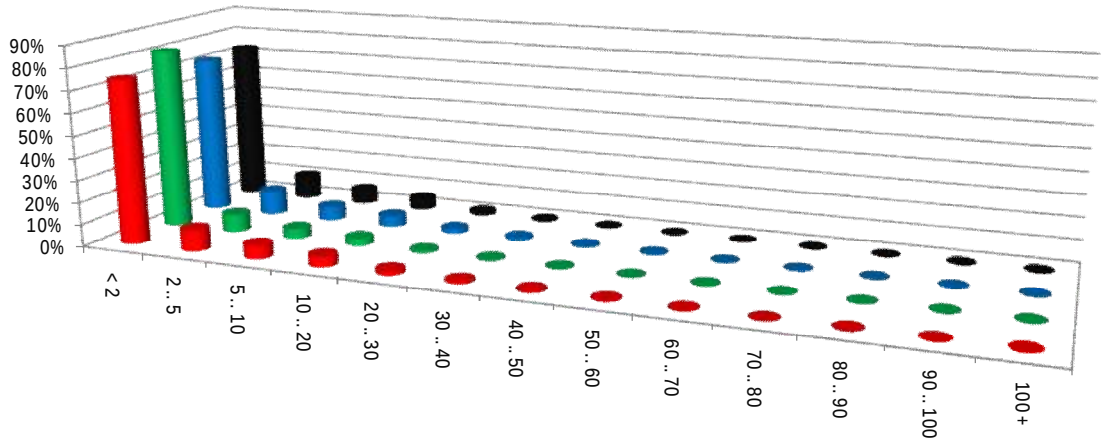
Як бачимо з таблиці 2, відображення отриманих результатів, алгоритм RS-аналізу не виявляє ЦВЗ у зображеннях.

Дослідження впливу на результати RS-CA масок різного розміру

Розробники методу RS-CA стверджують [6], що можливе використання масок будь-якого формату, а 2x2 наведено (рис. 5, 6), як варіант (приклад).

Наше дослідження методу RS-CA з масками різних форматів показало, що скануючі вікна з непарними розмірами не дають вірного результату, також комбінації з непарних і парних чисел до розміру 4x3 також призводять до великої помилки у результатах. Тобто маски 1x1, 1x2, 1x3, 2x1, 2x3, 3x1, 3x2, 3x3, 5x5, 7x7, 9x9, 11x11, 13x13, 15x15, 17x17, 19x19 – призводять до чисельно невірного результату, що виражається у великому відсотку ХПВС для пустих контейнерів (від 50 до понад 100, причому часто трапляються значення 200, 300, 400 %). На рис. 7 представлено результати для маски 3x3 для вибірки з 53306 фотографій. Видно, що кількість аномальних значень для такого вікна є неприйнятною. Головним чинником зміщення максимуму гістограми на позицію 10-40 % є рівень ХПВС у блакитному каналі, для якого характерне сильне зашумлення. Підвищену чутливість RS-CA з непарними масками до природних шумів можна використати в задачах цифрової обробки зображень.

Навпаки, використання вікон розмірів: 1x4, 2x4, 4x1, 4x2, 4x3, 4x4, 6x6 дає значення ХПВС аналогічні масці 2x2 з точністю менше 1 відсотку, тобто в межах чисельних похибок округлення.



	< 2	2 .. 5	5 .. 10	10 .. 20	20 .. 30	30 .. 40	40 .. 50	50 .. 60	60 .. 70	70 .. 80	80 .. 90	90 .. 100	100 +
Червоний	75%	10%	6%	5%	2%	1%	1%	1%	0%	0%	0%	0%	0%
Зелений	82%	8%	5%	3%	1%	1%	0%	0%	0%	0%	0%	0%	0%
Блакитний	72%	11%	7%	5%	2%	1%	1%	0%	0%	0%	0%	0%	0%
Всереднений	73,40%	11,07%	6,97%	5,42%	1,75%	0,64%	0,27%	0,14%	0,05%	0,04%	0,06%	0,05%	0,14%

Рис. 5. Гістограма ХПВС RS-CA для кольорових каналів, вікно 2x2

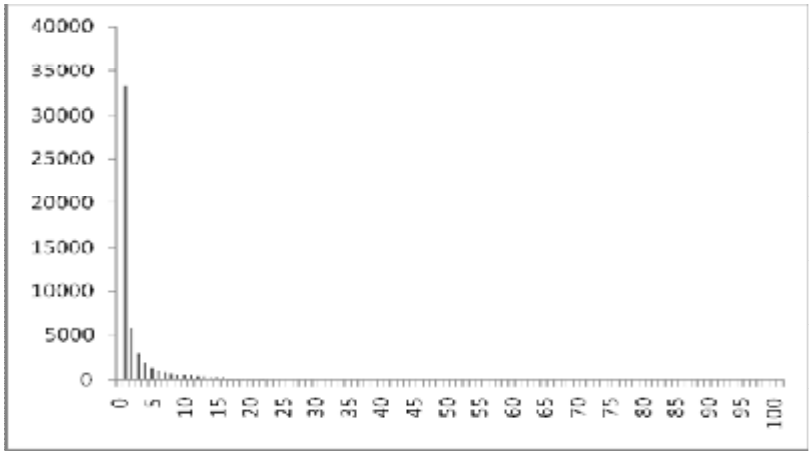
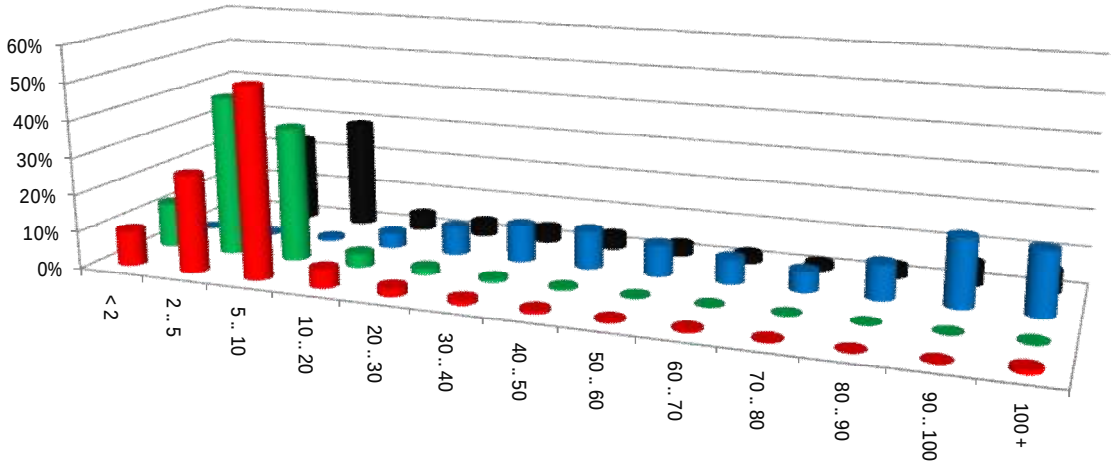


Рис. 6. Загальна гістограма (спектр) RS-CA для вибірки з 53306 файлів, маска 2x2

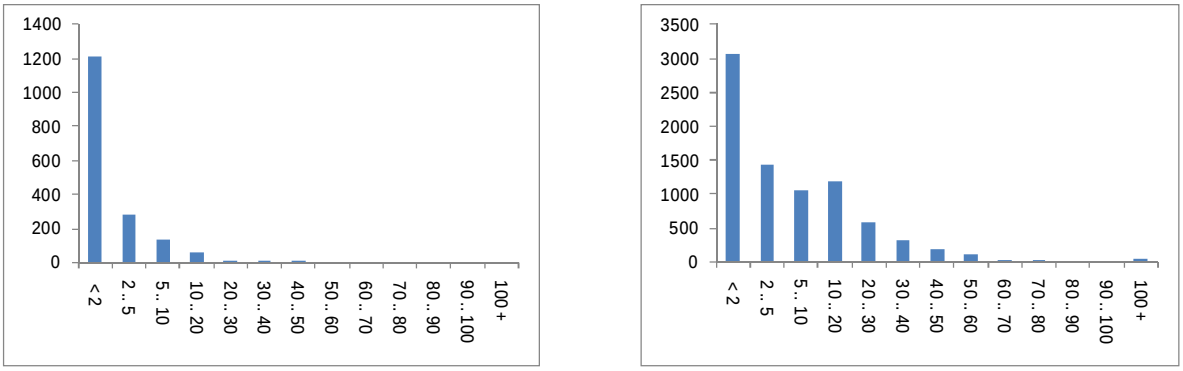


	< 2	2 .. 5	5 .. 10	10 .. 20	20 .. 30	30 .. 40	40 .. 50	50 .. 60	60 .. 70	70 .. 80	80 .. 90	90 .. 100	100 +
Червоний	10%	27%	52%	5%	2%	1%	1%	1%	1%	0%	0%	0%	1%
Зелений	12%	43%	36%	4%	2%	1%	1%	0%	0%	0%	0%	0%	0%
Блакитний	0%	0%	1%	4%	8%	10%	11%	9%	7%	6%	10%	18%	17%
Всереднений	7%	23%	30%	4%	4%	4%	4%	3%	3%	2%	3%	6%	6%

Рис. 7. Гістограми ХВПС, вікно 3x3 для кольорових каналів вибірки і осереднена по каналах

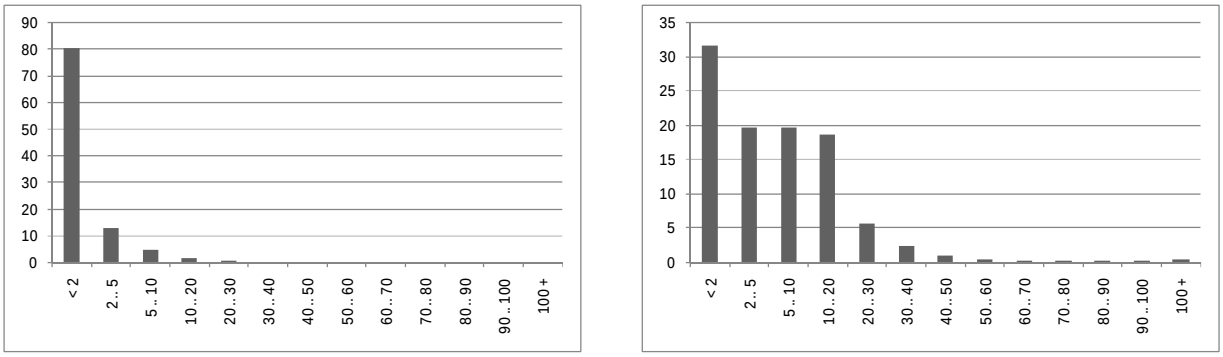
Оскільки час обчислення на платформі Intel i7 920 збільшується до 1,5 години для 10 файлів на перелічених форматах скануючого вікна дослідження на великих вибірках не виконувались. Нашою рекомендацією є застосовувати методу RS-CA з вікнами розміром 2x2, оскільки такий формат надає найбільшу точність виявлення прихованих даних при максимальній продуктивності обчислень.

При визначенні порогового значення для виявлення прихованих повідомлень слід також враховувати вік цифрової фотографії. Наприклад, для цифрових фотографій, зроблених 10-15 років тому, характерний більш високий рівень ХПВС для всіх кольорових каналів та більш повільне спадання в діапазоні від 5-30 %, порівняно з фотографіями, зробленими за останні 2-3 роки рис. 8-9. Тому поріг рівень виявлення прихованих даних у 2 % ХПВС не прийнятний для фотографій зроблених 10-15 років тому, для яких поріг складає 10-30 % залежно від якості ПЗЗ-матриці пристрою реєстрації і параметрів експозиції.



а **б**

Рис. 8. Гістограми ХПВС блакитного каналу, вікно 2x2 для нових цифрових фотографій (а) та фотографій, зроблених 10-15 років тому (б)



а **б**

Рис. 9. Гістограми всередених величин ХПВС, вікно 2x2 для нових цифрових фотографій (а) та фотографій, зроблених 10-15 років тому (б)

Реалізація RS-CA на архітектурі CUDA

Як зазначалось вище, для підтвердження статистичної достовірності гіпотез необхідно розраховувати статистики на великих масивах зображень, що складають десятки тисяч файлів. При цьому час обробки одного файлу залежить від його розміру та швидкодії процесору і для великих файлів може складати десятки секунд, навіть для сучасних процесорів типу Intel Core i7. Час обробки всієї вибірки може сягати декількох діб; наприклад якщо при обробці одного файлу, час обробки становитиме 10 с то обробка 10000 файлів займе 27 годин. Обробка вибірок більшого розміру чи вибірок, що містять файли більшого розміру, займе декілька діб. В процесі досліджень може виникати велика кількість гіпотез, які необхідно перевіряти, що в умовах відсутності доступних виділених обчислювальних ресурсів, займає багато часу і обмежує просування досліджень.

Час обробки зменшується при використанні багатоядерних процесорів, проте таке зменшення обмежується кількістю ядер процесора, яка в сучасних процесорах становить 2-6. Для подальшого пришвидшення обробки зображень методом RS-CA (та іншими) необхідно шукати альтернативні варіанти реалізації цих методів.

Так, одним з найбільш перспективних напрямків є реалізація методів аналізу зображень під архітектуру CUDA (Compute Unified Device Architecture). Вона реалізована в графічних процесорах фірми NVidia та передбачає написання програм, що будуть виконуватись на векторних мультипроцесорах, що містять по декілька ядер; сумарно в одному графічному процесорі може міститись декілька сотень обчислювальних ядер, що відкриває значно більші потенційні можливості для прискорення обчислень, які необхідні для розв'язання певних задач. Звичайно, ці обчислювальні ядра не еквівалентні ядрам центрального процесора, є певні обмеження, які необхідно врахувати для їх ефективного використання.

Одне з основних обмежень – робота з пам'яттю, яка є загальним ресурсом для ядер графічного процесора.

На даний час авторами статті реалізована попередня версія модуля-розширення комплексу стегоаналітичних досліджень, який реалізує метод RS-CA під архітектуру CUDA. Процес оптимізації алгоритму ще не завершено, проте вже зараз ця реалізація працює в 15-20 разів швидше, ніж однопоточна версія, що виконується на досить потужному центральному процесорі. Це підтверджує доцільність використання архітектури CUDA для реалізації методів обробки та аналізу зображень, роботи в даному напрямку будуть продовжені.

Висновки

1. Вперше запропоновано комплексну систему статистичних досліджень для методів стегоаналізу.
2. Розроблено новий модульний проблемно-орієнтований комплекс стегоаналітичних досліджень, що дозволяє оператору гнучко змінювати напрямки збору статистики та експортувати отримані дані в додаток Майкрософт Ексель (електроні таблиці).
3. Завдяки створеному комплексу отримано нові результати в області приховування даних стеганографічними методами у зображеннях:

- виявлено обмеження методу RS-стегоаналізу для багатьох класів зображень, доступних у мережі Інтернет;
- експериментально доведено, що раціональний розмір скануючого вікна – 2x2 та спростовано твердження про використання масок будь-якого формату.
- запропоновано способи обходу методу RS-стегоаналізу;
- показано, що для великих масивів зображень характерні залежності між статистичними характеристиками кольорових каналів з декількома максимумами;
- на основі статистичних досліджень показано, що RS-CA неефективний у виявленні ЦВЗ;
- виявлено, що максимально допустимий об'єм даних безпечного вбудовування даних складає 2 %, а задовільний рівень не повинен перевищувати 5 %. Об'єми вбудовування 5-10 % є підозрілими, а понад 10 % малоймовірними для необроблених цифрових фотографій.

4. Перспективними напрямками є дослідження нового методу WS-стегоаналізу та побудова аналогічного комплексу для стеганографії зображень у форматі JPEG, дослідження змін співвідношень між характеристиками кольорових каналів після фільтрації зображень та застосування методів ІАД до структурованої колекції зображень.
5. Реалізація методів обробки та аналізу зображень під архітектуру CUDA може пришвидшити обробку зображень в багато разів, що суттєво прискорює проведення стегоаналітичних досліджень.

Література

1. Задирака В. К. Комп'ютерна стеганографія : [монографія] / Серієнко І. В., Коваленко І. М., Андон П. І. та ін. // Стан та перспективи розвитку інформатики в Україні – К. : Наук. думка, 2010. – С. 736–747.
2. Конахович Г.Ф. Компьютерная стеганография. Теория и практика / Г. Ф. Конахович, А. Ю. Пузыренко. – К. : МК-Пресс, 2006. – 288 с.
3. Advanced Statistical Steganalysis / R. Böhme, Springer, 2010.
4. Steganography in Digital Media: Principles, Algorithms, and Applications / J. Fridrich, Cambrige University Press, 2010.
5. Стеганография, цифровые водяные знаки и стеганоанализ : [монография] / А. В. Аграновский, А. В. Балакин, В. Г. Грибунин, С. А. Сапожников. – М. : Вузовская книга, 2009. – 220 с.
6. J.Fridrich, M. Goljan, and R. Du, "Detecting LSB Steganography in Color and Gray-Scale Images", Magazine of IEEE Multimedia, Special Issue on Security, October-November issue, 2001, pp. 22– 28.
7. Королёв В. Ю. Стеганография по методу наименее значимого бита на базе персонализированных флеш-накопителей / В. Ю. Королёв, В. В. Полиновский, В. А. Герасименко // Управляющие системы и машины. – № 1 (231). – 2011. – С. 79–87.
8. Корольов В. Ю. RS-стеганоаналіз. Принципи роботи, недоліки та концепція метода його обходу / В. Ю. Корольов, В. В. Поліновський, В. А. Герасименко // Вісник Вінницького політехнічного інституту. – 2010. – № 6. – С. 66–71.
9. Корольов В. Ю. Визначення можливостей RS-стегоаналізу для дослідження статистичних властивостей зображень / В. Ю. Корольов, В. В. Поліновський, В. А. Герасименко // Вісник Хмельницького національного університету. – 2010. – № 4. – С. 102–110.
10. Корольов В. Ю. Стеганографічна персоналізація інформації на базі ПК / В. Ю. Корольов, В. В. Поліновський, В. А. Герасименко // Вісті Академії інженерних наук України. – 2009. – № 2 (39). – С. 18– 24.
11. Fridrich J., Goljan M., Du R. Lossless Data Embedding – New Paradigm in Digital Watermarking // Special Issue on Emerging Applications of Multimedia Data Hiding. – 2002. – Vol. 2002, N 2. – P. 185– 196.
12. S.Geetha, N.Kamaraj Optimized Image Steganalysis through Feature Selection using MBEGA // International Journal of Computer Networks & Communications (IJCNC), Vol.2, № 4, July 2010, P.161–175. http://128.84.158.119/ftp/arxiv/papers/1008/1008.2824.pdf
13. http://diit.sourceforge.net/
14. http://dde.binghamton.edu/, http://dde.binghamton.edu/download/

15. http://vsl.sourceforge.net/
16. http://opds.sut.ru/electronic_manuals/pe/

Надійшла 24.8.2011 р.

УДК 621.376.6

А.А. ОВЧАРУК, С.Т. БАРАСЬ
Вінницький національний технічний університет

ОПТИМІЗАЦІЯ АЛГОРИТМУ КВАДРАТУРНОЇ АМПЛІТУДНОЇ МОДУЛЯЦІЇ

Стисло розглянуто алгоритм квадратурної амплітудної модуляції. Запропоновано метод підвищення швидкості передачі інформації на основі алгоритму квадратурної амплітудної модуляції за рахунок введення блоку перекомутації частот-носіїв у структуру передавача. Проведено оцінку можливого приросту швидкості передачі інформації, який виникає в результаті введення блоку перекомутації частот-носіїв.

The algorithm of the quadrature amplitude modulation is shortly considered. The method of increasing the data transmission rate based on the algorithm of the quadrature amplitude modulation by introducing a block of carrier frequencies overcommutation in transmitter structure is proposed. The estimation of potential growth rate of information transfer, arising from the introducing a block of carrier frequencies overcommutation is conduct.

Ключові слова: телекомунікації, квадратурна амплітудна модуляція, КАМ, швидкість передачі інформації.

Вступ

Швидкість передачі інформації є одним із основних параметрів сучасних цифрових систем зв'язку. Висока швидкість передавання досягається різними шляхами, одним з яких є використання алгоритму квадратурної амплітудної модуляції (КАМ).

У алгоритмі КАМ використовується два інформаційних параметри сигналу: початкова фаза і амплітуда. Традиційним підходом для підвищення швидкості передачі інформації на основі використання КАМ вважається збільшення кількості рівнів існуючих інформаційних параметрів та встановлення такого співвідношення сигнал/шум, при якому кількість помилок є допустимою [1].

Враховуючи те, що збільшення рівнів сигналу призводить до зростання міжрівневих спотворень, а, отже, і до збільшення кількості помилок, можна запропонувати ще один підхід по підвищенню швидкості передачі інформації на основі КАМ без суттєвого збільшення кількості помилок. При цьому передбачається введення ще одного інформаційного параметра – миттєвої фази сигналу, зміна якої забезпечується перекомутацією частот-носіїв під час існування окремого імпульсу модулюючого сигналу.

Постановка завдання

Метою даного дослідження є підвищення швидкості передачі інформації на основі алгоритму КАМ за рахунок використання миттєвої фази сигналу під час існування окремого імпульсу модулюючого сигналу як додаткового інформаційного параметру, а також визначення можливого приросту швидкості передачі інформації в результаті введення додаткового інформаційного параметра.

Методика проведення досліджень

При використанні алгоритму КАМ сигнал, що передається, створюється одночасними змінами амплітуди синфазної (I) і квадратурної (Q) компонент несучого гармонійного коливання (f_c), які зміщені по фазі одна відносно одної на $p/2$. Результуючий сигнал Z являє собою суму цих складових [1-3]. Таким чином, дискретний сигнал з КАМ може бути представлений співвідношенням:

$$Z_m(t) = I_m \cdot \cos(2pf_c t) + Q_m \cdot \sin(2pf_c t), \quad (1)$$

де t – змінюється в діапазоні $\{(m-1)\cdot\Delta t...m\cdot\Delta t\}$;

m – порядковий номер дискрету часу модулюючого сигналу;

Δt – крок квантування модулюючого сигналу за часом.

Отже у алгоритмі КАМ передбачається використання двох паралельних амплітудних модуляторів на які подаються частоти-носії, зміщені по фазі одна відносно одної на $p/2$. Спрощена структурна схема квадратурного модулятора подана на рис. 1.

Значення I_m та Q_m визначаються за формулами:

$$\begin{aligned} I_m &= a_m \cdot p; \\ Q_m &= b_m \cdot p, \end{aligned} \quad (2)$$

де a_m і b_m – модуляційні коефіцієнти;

p – крок квантування модулюючого сигналу по амплітуді.

Значення модуляційних a_m і b_m для алгоритму КАМ-4 надані у таблиці 1.