DOI: https://doi.org/10.15276/ict.02.2025.10
УДК 004.056.5:003.27

Реалізація та порівняльний аналіз методів текстової стеганографії

Хом’як Роман Миколайович[1)]
Магістр каф. Безпеки інформаційних технологій
ORCID: https://orcid.org/0009-0007-0837-8402; roman.khomiak.mkbbi.2024@lpnu.ua
Мельник Віктор Анатолійович[1)]
Д-р техніч. наук, професор каф. Безпеки інформаційних технологій
ORCID: https://orcid.org/0000-0002-5687-4204; viktor.a.melnyk@lpnu.ua
[1)] Національний університет “Львівська політехніка”, вул. Степана Бандери, 12. Львів, 79000, Україна

АНОТАЦІЯ

Стеганографія – важливий напрям сучасної інформаційної безпеки, що приховує дані у носії таким чином, щоб сам факт секретної передачі залишався непомітним. Серед різних її видів текстова стеганографія займає особливе місце через поширеність текстового формату, хоча й має суттєве обмеження – малу ємність контейнера порівняно з аудіо чи зображеннями. На практиці текстова стеганографія відіграє ключову роль у захисті інтелектуальної власності, виступаючи як «невидимий соціальний DRM» (Digital Right Management): в кожен примірник електронної книги можна вбудувати унікальний непомітний ідентифікатор (маркер), який не змінює змісту для читача, але дає змогу ефективно відстежувати джерела витоку або несанкціонованого поширення контенту.

З огляду на об'єктивну обмеженість обсягу тексту як контейнера для прихованої інформації, в даній статті приділено особливу увагу ретельному порівнянню й оцінці методів. **Мета полягала** у виборі підходів, що дозволяють максимально ефективно приховувати дані, зберігаючи при цьому первісну структуру та зовнішній вигляд документа.

У ході ґрунтовного дослідження було обрано та практично реалізовано два методи:

– метод заміни символів однакового накреслення (використання різних варіантів одного й того ж знака, які візуально не відрізняються для читача, але кодують приховані дані);

– метод зміни порядку проходження знаків кінця рядка (де варіації в оформленні абзаців або рядків непомітно несуть у собі закодовану інформацію).

Отримані результати порівняльного аналізу переконливо показують, що текстова стеганографія, незважаючи на об'єктивне обмеження малою ємністю контейнера, є перспективним і надійним інструментом для індивідуалізації та захисту електронних публікацій. Застосування таких методів, як заміна знаків однакового накреслення та зміна порядку проходження знаків кінця рядка, забезпечує високий рівень непомітності прихованого маркеру для кінцевого читача. Одночасно досягається стійкість вбудованого ідентифікатора до простих маніпуляцій і копіювання.

Така комбінація характеристик дозволяє ефективно масштабувати рішення у сфері цифрових прав (соціальний DRM), сприяючи формуванню нових, захищених підходів до поширення чутливого інтелектуального контенту. Для практичної демонстрації цих переваг було створено демонстраційний програмний засіб, який, використовуючи обидва згадані стеганографічні методи у поєднанні з додатковим шифруванням, наочно підтверджує можливість поєднання непомітності, стійкості та ефективності.

Ключові слова: текстова стеганографія; контейнер тексту; захист вмісту; метод заміни символів однакового накреслення; метод заміни знаків кінця рядка

Актуальність. Основною задачею стеганографії є приховування самого факту існування та передачі таємного повідомлення. Цим вона принципово відрізняється від криптографії, яка лише шифрує вміст, роблячи його незрозумілим для зловмисника, але не приховує факт його існування.

У сучасному світі стеганографічні системи вирішують низку важливих завдань [1], зокрема:

– захист інформації від несанкціонованого доступу;
– захист авторських прав;
– подолання систем моніторингу мережевого трафіку;
– приховування програмного забезпечення;
– створення прихованих каналів витоку інформації.

Текстова стеганографія є актуальною через зростання обсягів електронного обміну даними та потребу в прихованих каналах комунікації. Вона дозволяє вбудовувати секретну інформацію у звичайний текст без зміни його змісту, що важливо не лише для захисту

інтелектуальної власності, а й для забезпечення конфіденційності, автентифікації повідомлень, створення унікальних ідентифікаторів або цифрових водяних знаків. Обмеженість текстового контейнера стимулює пошук нових методів і алгоритмів, здатних поєднати непомітність, стійкість і достатню ємність.

Метою дослідження є аналіз та пошук ефективних сучасних методів текстової стеганографії, а також реалізація обраних методів в рамках демонстрації принципів роботи алгоритмів та порівняння і оцінювання їх характеристик. В основі дослідження лежить порівняння існуючих підходів до приховування інформації у текстових даних, їх оцінка за критеріями ефективності використання контейнера, непомітності та стійкості до форматування. Результати дослідження можуть слугувати основою для подальшої розробки оптимізованих рішень у сфері текстової стеганографії, які відповідатимуть сучасним викликам інформаційної безпеки та цифрового контент-менеджменту.

Аналіз досліджень і публікацій. Стеганографія, як мистецтво приховування таємних повідомлень, існує ще з давніх часів. У минулому використовувалися матеріальні контейнери, як-от воскові дощечки або листи, і такі методи належать до так званої "традиційної" стеганографії.

Із настанням ери комп’ютерів виник новий напрям – комп’ютерна стеганографія [2]. Вона охоплює методи, що ґрунтуються на особливостях форматів файлів, файлових систем та протоколів. Деякі з цих методів використовують зарезервовані поля файлів, що рідко або взагалі не застосовуються, інші – властивості, які не відображаються на екрані. До цього напряму належать і методи текстової стеганографії [3].

Водночас розвивається й цифрова стеганографія [2], яка оперує з контейнерами аналогової природи, що мають надлишок шумових сигналів. Незначні зміни в таких файлах залишаються непомітними для сприйняття людиною завдяки особливостям її зорової та слухової систем. До контейнерів цифрової стеганографії відносяться файли фото, аудіо та відео. На відміну від них, текстові методи приховування інформації мають у своєму розпорядженні доволі обмежений обсяг надлишкових даних.

Методи текстової стеганографії поділяються на три головні категорії [4]: методи форматування, лінгвістичні методи та методи статичної або випадкової генерації.

Методи форматування працюють з невидимими або малопомітними змінами в структурі самого текстового файлу, наприклад, змінюючи пробіли, відступи, регістр літер, шрифти, інтервал між словами чи рядками, не змінюючи при цьому фактичного тексту [4].

Лінгвістичні методи використовують природну мову для приховування інформації, замінюючи синоніми, змінюючи порядок слів чи речень або вбудовуючи повідомлення через контекстно-вільні граматики, створюючи при цьому синтаксично коректний і змістовний маскувальний текст [3].

Методи статичної або випадкової генерації створюють абсолютно новий маскувальний текст, що не має попередника, використовуючи, наприклад, ланцюги Маркова, або ж сучасні Великі Мовні Моделі (LLM), які генерують текст, що виглядає природно, але містить приховане повідомлення [5].

Постановка задачі. Одним з перспективних напрямів застосування стеганографії є захист інтелектуальної власності, в особливості електронних книг. Наразі компанії використовують системи DRM (Digital Right Management), сенс яких – обмежити користувача в правах на використання книги: копіювання, друк, поширення, передача файлу і т.д. Проте існують шляхи обходу даних систем та поширення ліцензованих копій стає важко відслідковувати. На заміну приходить соціальний DRM – використання видимих та невидимих цифрових водяних знаків, що включають інформацію про покупця, транзакцію, та сторону, що надала книгу. Видимі цифрові водяні знаки можна з легкістю очистити так само, як і невидимі, що часто вставляють у файли форматування, що легко очищаються [6]. Проте, якщо використати DRM разом з стеганографією, яка не буде залежати від форматування файлу – то стає набагато простіше вирахувати того, хто став джерелом піратської копії.

Основною задачею цієї роботи буде визначення такого методу текстової стеганографії, що буде ефективно використовувати текст як контейнер та не буде вразливим до змін шрифтів, регістру літер, та не змінюватиме сенс та контекст оригінального тексту.

Огляд методів текстової стеганографії. Основою для класифікації методів стеганографії є контейнери, в яких приховується інформація. Вибір контейнера та спосіб його модифікації визначають стійкість та пропускну здатність методу.

Оскільки головним контейнером в нашому дослідженні є сам текст, то методи можна розділити на такі категорії:

– методи перекручування формату тексту, які використовують пробіли, відступи, розриви рядків та інші малопомітні символи для кодування повідомлень;

– синтаксичні методи, що змінюють пунктуацію та структуру речень у межах граматичних правил;

– семантичні методи, де для приховування даних використовують синонімічні заміни слів чи стилістичні перефразування;

– інші методи, які поєднують кілька підходів або використовують нестандартні властивості текстових форматів.

Методи перекручування формату файлів найчастіше базуються на використанні пробілів або непомітних символів. Один із найпростіших з них – метод додаткових пробілів у кінці речень. Він дозволяє кодувати інформацію за рахунок того, що на завершенні речення додається один або два пробіли. Якщо використовується один пробіл - це може відповідати значенню “0”, а два пробіли – «1». Інший варіант - метод хвостових пробілів, який працює на рівні рядків: у кінець кожного рядка додається певна кількість пробілів, що відповідає певній послідовності бітів. У класичному варіанті цей метод кодує 1 біт на рядок, проте його вдосконалена модифікація дозволяє кодувати до пів-байта повідомлення в одному рядку, використовуючи різні комбінації пробілів і табуляцій. Деякі дослідження також пропонують застосовувати невидимі символи Unicode, наприклад “нульову ширину пробілу” або “нульову ширину приєднання”, що значно ускладнює виявлення прихованих даних візуально [7]. Однак усі ці методи залишаються вразливими, оскільки сучасні текстові редактори чи системи форматування можуть автоматично видаляти або нормалізувати надлишкові пробіли та непомітні символи, внаслідок чого приховане повідомлення буде втрачено.

Синтаксичні методи кодування інформації ґрунтуються на змінах пунктуації, використанні різних варіантів розділових знаків або перебудові граматичної структури. Вони застосовуються у випадках, коли існує кілька рівноцінних з погляду граматики варіантів оформлення тексту. Наприклад, у реченні “Контейнер, стеганоключ і стеганограма” кома може бути відсутня перед останнім елементом, що відповідає значенню “1”, тоді як варіант із комою “Контейнер, стеганоключ, стеганограма” може відповідати значенню “0”. У більш складних випадках використовуються різні варіанти лапок ("...", “...”, ‘...’), які можуть кодувати кілька бітів. Додатково можна варіювати розташування тире і дефісів, використання подвійних або одинарних пропусків після крапки, а також вибір між різними варіантами структурних роздільників у текстах програмного чи технічного характеру. Ще один підхід полягає у використанні синонімічних пунктуаційних конструкцій, наприклад: “однак” може виділятися комами або тире, і вибір варіанта може відповідати прихованому біту. Попри різноманітність способів, ефективність використання контейнера у таких методів низька [1].

Семантичні методи стеганографії працюють на рівні змісту й засновані на заміні слів на їхні синоніми, використанні парафразів або стилістичних варіацій. Наприклад, слова “швидкий” та “прудкий” можуть позначати різні значення в прихованому коді. Кількість інформації, що може бути закодована, залежить від числа доступних синонімів для конкретного слова. Якщо для слова існує лише одна-дві альтернативи, прихований канал має низьку пропускну здатність, а якщо варіантів значно більше – передавати інформацію стає ефективніше. Окремим підвидом є метод тематичних замін, коли у технічних текстах використовують різні рівнозначні терміни (“файл” / “документ”, “сервер” / “хост”).

Складність полягає у збереженні природності тексту: некоректна заміна може виглядати неприродно для читача й викликати підозру [1].

Окрім основних методів, існують й інші підходи, які не потрапляють у класичні категорії. Один із таких – метод зміни порядку проходження маркерів кінця рядка (CR і LF). У різних операційних системах застосовуються різні позначення кінця рядка: у Windows – CRLF, у Unix-подібних системах – LF, а в старих Mac – CR. Заміна одного варіанта на інший або їх чергування може кодувати один біт інформації на кожній заміні. Хоча пропускна здатність такого підходу невелика, він відрізняється підвищеною стійкістю, оскільки більшість редакторів не змінюють формат кінця рядка автоматично. Ще один цікавий метод – заміна літер латиниці на кириличні літери з однаковим накресленням та навпаки (наприклад а - a, Х - X, р - p і т.д.). Така підміна практично непомітна для людини, але дозволяє приховувати великі обсяги даних у довгих текстах. Схожий принцип використовується і в методі зміни регістру символів: один варіант (“Слово”) може відповідати ‘0’, інший (“слово”) – ‘1’. Цей спосіб має найбільшу пропускну здатність серед простих методів, оскільки кодування можливе для кожного символу, за винятком пробілу. Проте він легко виявляється автоматичними аналізаторами, особливо у випадках, коли текст не містить виправданих змін регістру. Для ускладнення виявлення іноді застосовуються спеціальні шрифти, де велика й мала літера виглядають однаково [7].

Аналіз характеристик методів текстової стеганографії. Вибір методів текстової стеганографії для практичної реалізації залежить від багатьох аспектів. Насамперед враховується ефективність використання контейнера, оскільки саме вона визначає, який обсяг прихованих даних можна вбудувати у текст без суттєвих змін його структури. Якщо метод має низьку пропускну здатність, його використання є недоцільним.

Другим критичним фактором є стійкість до стегоаналізу. Будь-яка зміна у тексті не повинна викликати підозри у читача чи автоматизованої системи перевірки. Це означає, що метод має забезпечувати мінімальне або зовсім непомітне спотворення тексту, зберігаючи природність мови й форматування.

Не менш важливою характеристикою є стійкість до спотворень. Приховане повідомлення не повинно втрачатися внаслідок технічних змін, наприклад, копіювання тексту, зміни кодування чи застосування іншого шрифту. Якщо інформація зникає після відкриття документа у новому редакторі або після автоматичного форматування, метод втрачає практичну цінність.

Останній, але важливий параметр - складність реалізації. Деякі методи прості й не вимагають спеціалізованих бібліотек, тоді як інші потребують складних механізмів (аналізу тексту, словників чи машинного навчання). Метод, що поєднує низьку ефективність з високою складністю реалізації, є недоцільним для застосування.

Таким чином, оптимальний вибір методу досягається лише шляхом компромісу між чотирма ключовими характеристиками: пропускною здатністю, стійкістю до спотворень, стійкістю до стегоаналізу та складністю реалізації. Для наочності складемо порівняльну характеристику основних методів текстової стеганографії (табл. 1).

Для реалізації розробленої авторами в межах цієї роботи демонстраційної програми було обрано методи зміни порядку проходження маркерів кінця рядка та метод знаків однакового накреслення.

Розроблення алгоритмів. Перед розробкою програмного засобу потрібно розробити алгоритм, за яким буде працювати програма та алгоритми реалізації вибраних стеганографічних методів.

На рис. 1 наведено два алгоритми роботи програми. Перший показує use-case приховування повідомлення в контейнер. Другий навпаки, показує use-case видобування повідомлення з контейнера.

Таблиця 1. **Порівняльна характеристика методів текстової стеганографії**

Категорії методів	Характеристика / Назва методу	Відносна пропускна здатність	Стійкість до спотворень	Стійкість до стегоаналізу	Складність реалізації
Методи перекручування формату	Метод зміни регістру	В	Н	Н	Н
	Метод додаткових пробілів у кінці речення	Н	Н	Н	Н
	Метод хвостових пробілів	Н	Н	С	Н
	Модифікований метод хвостових пробілів	С	Н	С	Н
	Використання невидимих символів Unicode	С	Н	С	Н
Синтаксичні методи	Варіанти пунктуації, лапки, тире.	Н	С	Н	В
Семантичні методи	Синоніми, парафрази, тематичні заміни	Н	С	С	В
Інші методи	Метод зміни порядку проходження маркерів кінця рядка (CR/LF)	Н	С	В	С
	Метод знаків однакового накреслення	С	С	С	С

Примітка: Оцінювання виконувалось за відносною шкалою:
Н – низька
С – середня
В – висока

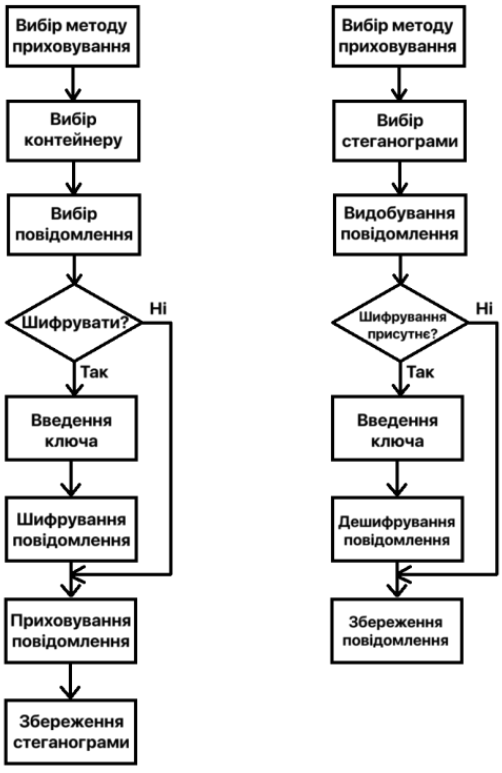

Рис. 1. **Основні use-case роботи програми**

Як видно з блок-схеми, в програмі додатково введено функцію шифрування для додаткового шару захисту. За алгоритм шифрування було обрано 3-DES, оскільки він цілком задовольняє нашим цілям. Введений пароль буде гешовано за допомогою SHA-256.

Перед написанням алгоритмів реалізації стеганографічних методів потрібно розібратись із специфікою використання контейнера в даних методах. В методі зміни порядку проходження кінця рядка використовується логіка наведена в табл. 2. Для методу знаків однакового накреслення було підготовлено алфавіт для заміни знаків з однаковим накресленням. Наведений він в табл. 3.

Таблиця 2. **Специфіка приховування повідомлення методом зміни порядку проходження знаків кінця рядка**

Біт повідомлення	Порядок маркерів
“1”	LF/CR
“0”	CR/LF

Таблиця 3. **Специфіка приховування повідомлення методом знаків однакового накреслення**

	Розкладка / Hex код символу ASCII			
	Латиниця	Кирилиця	Латиниця	Кирилиця
Літери	A (0x41)	А (0xC0)	T (0x54)	Т (0xD2)
	B (0x42)	В (0xC2)	X (0x58)	Х (0xD5)
	C (0x43)	С (0xD1)	a (0x61)	а (0xE0)
	E (0x45)	Е (0xC5)	c (0x63)	с (0xF1)
	H (0x48)	Н (0xCD)	e (0x65)	е (0xE5)
	I (0x49)	І (0xB2)	i (0x69)	і (0xB3)
	K (0x4B)	К (0xCA)	o (0x6F)	о (0xEE)
	M (0x4D)	М (0xCC)	p (0x70)	р (0xF0)
	O (0x4F)	О (0xCE)	x (0x78)	х (0xF5)
	P (0x50)	Р (0xD0)	y (0x79)	у (0xF3)

Використання текстового контейнера методом знаків однакового накреслення наведено на рис. 2, а методом зміни порядку проходження кінця рядку – на рис. 3.

Рис. 2. **Використання текстового контейнера методом знаків однакового накреслення**

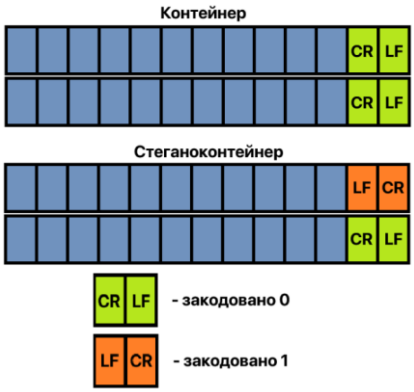

Рис. 3. **Використання текстового контейнера методом зміни порядку проходження знаків кінця рядка**

Додатково з повідомленням потрібно приховувати деяку службову інформацію, як наприклад, його заголовок. Для цього було вирішено приховувати: 1 біт як показник чи було повідомлення зашифровано, 31 біт як довжину самого повідомлення (максимальний розмір повідомлення складає 2047 мегабайтів) та 10 байтів на розширення повідомлення (для назв розширень до 10 символів включно). Схему цього заголовка наведено на рис. 4.

Рис. 4. **Загальна структура заголовка повідомлення**

Реалізація програмних засобів та експериментальні дослідження. За середовище розробки було обрано Visual Studio Code [8] через його широку підтримку різних платформ та інтуїтивний інтерфейс. Мовою розробки було обрано Python [9], оскільки вона має простий синтаксис та широкий вибір з бібліотек для роботи.

Серед використаних бібліотек: вбудований tkinter – для інтерфейсу, subprocess [10] – для переходу між підпрограмами, PyDES [11] – використана для алгоритму шифрування 3-DES, hashlib [12] – використана для гешування паролю користувача в ключ шифрування.

Головне вікно програми наведено на рис. 5.

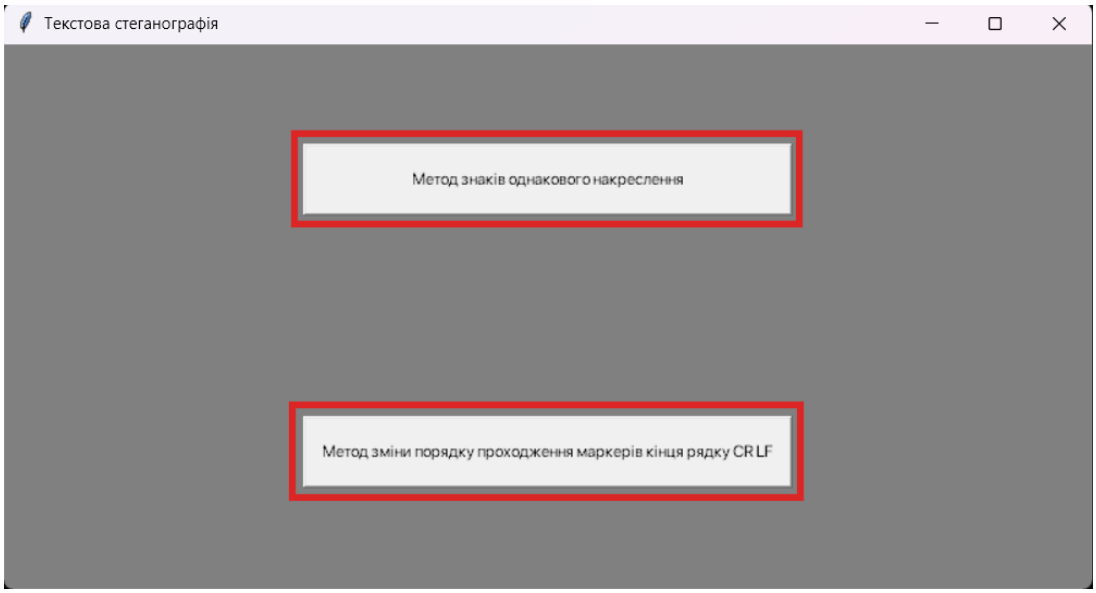

Рис. 5. **Головне меню програми**

При натисканні на відповідні кнопки меню програма відкриває вікно для роботи з відповідним методом. Вікна з відкритими методами продемонстровані на рис. 6.

За своєю структурою вікна не відрізняються. Робота розпочинається натисканням кнопки “Відкрити контейнер”, що дозволяє вибрати текстовий файл. Після вибору файлу його вміст буде виведено в текстових полях “Нех контейнер” та “Текст контейнера”. Максимально можливий об’єм повідомлення буде виведено в правій частині центрального блоку над кнопкою «Обрати повідомлення», як це показано на рис. 7.

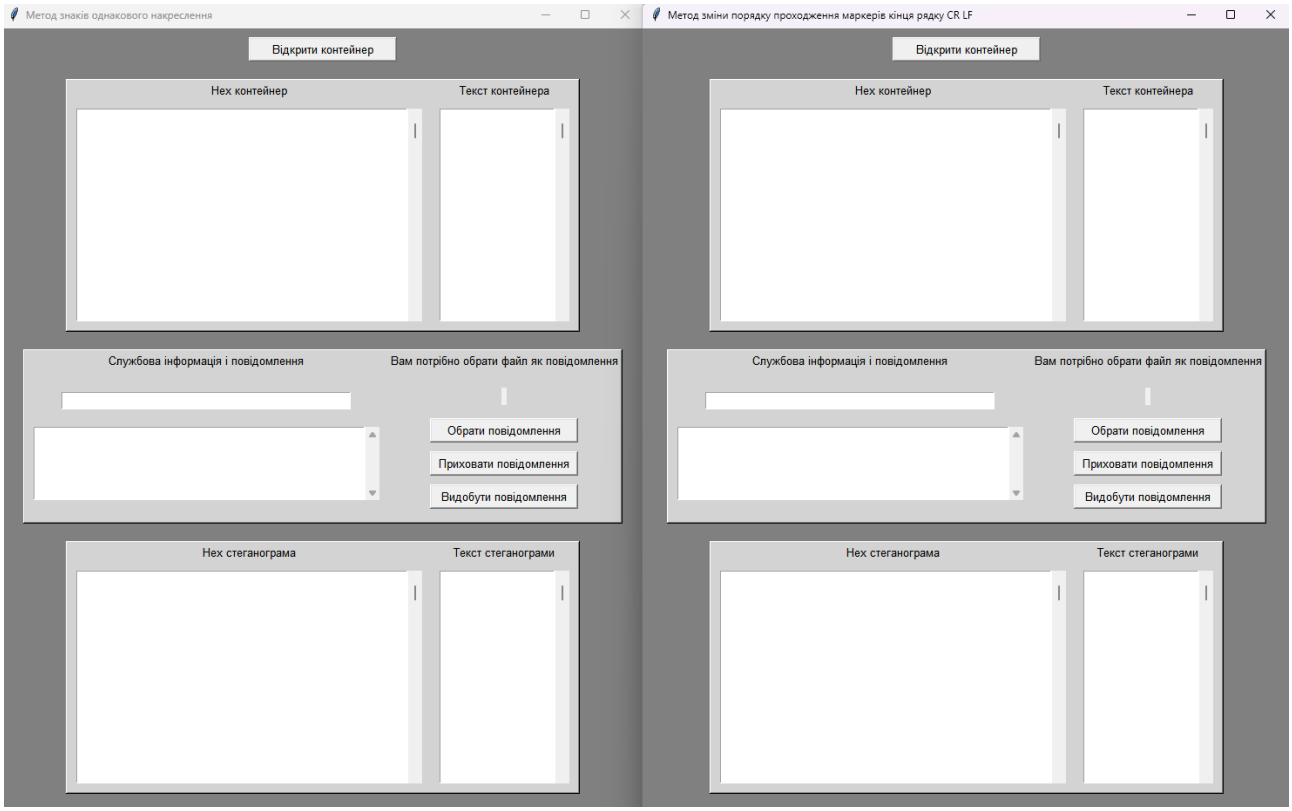

Рис. 6. **Вікна для роботи з обраними методами стеганографії**

Наступною операцією користувач повинен обрати повідомлення, що не виходить за межі зазначеного максимуму. Після цього потрібно обрати, чи потрібне додаткове шифрування. Якщо так – вводимо пароль і повідомлення буде зашифровано. Далі повідомлення буде виведено в ліву частину центральної панелі у вигляді Нех-коду разом зі службовою інформацією. Наступним кроком користувач натискає кнопку “Приховати повідомлення”, відразу після чого в поля нижньої панелі “Нех стеганограма” та “Текст стеганограми” виведеться готовий текст з прихованим повідомленням. Користувачу запропонують зберегти файл стеганограми.

Для видобування повідомлення слугує кнопка “Видобути повідомлення”. При натисканні на неї користувачеві потрібно обрати файл стеганограми, після чого, якщо повідомлення було шифроване, ввести пароль, що шляхом гешування перетвориться на ключ та розшифрує повідомлення. Отримане повідомлення буде запропоновано зберегти як файл.

Додатково в програмі присутній функціонал виділення повідомлення, що дозволяє порівняти контейнер та стеганограму і розглянути в яких місцях та які символи були замінені в методі знаків однакового накреслення та в яких місцях було змінено порядок знаків кінця рядка в методі зміни порядку проходження знаків кінця рядка.

Висновки: В роботі було проведено розгляд, аналіз та порівняння методів текстової стеганографії з метою вибору ефективних за декількома показниками: відносна пропускна здатність, стійкість до стегоаналізу, стійкість до спотворень та складність реалізації.

В ході аналізу було виявлено, що методи знаків однакового накреслення та зміни порядку проходження знаків кінця рядка виявились ефективнішими у порівнянні з іншими. В ході створення програмного забезпечення було розроблено алгоритми роботи застосунку та опрацьовано основну специфіку роботи методів з контейнерами.

Розроблена демонстраційна програма виконує приховування та видобування повідомлень в обрані контейнери, може додатково шифрувати повідомлення алгоритмом 3-DES та має функцію виділення, що показує, в яких місцях контейнера було приховано виділену частину повідомлення разом з такими ж місцями у стеганограмі для порівняння.

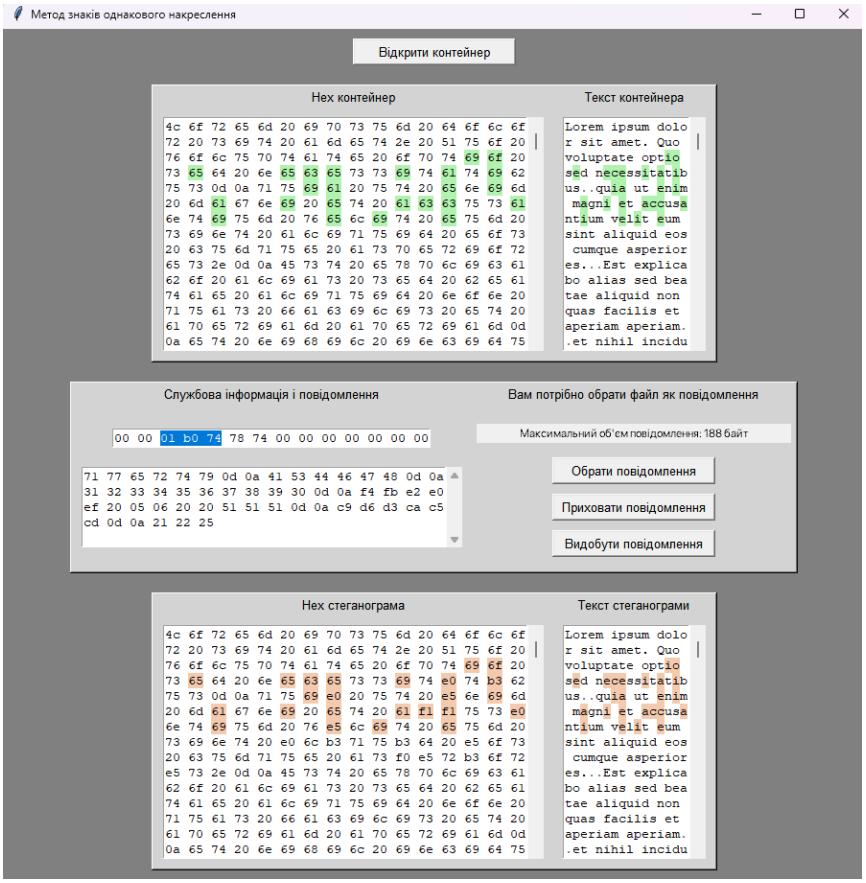

Рис. 7. **Процес приховування повідомлення за методом знаків однакового накреслення**

СПИСОК ЛІТЕРАТУРИ

1. Хорошко В. О., Яремчук Ю. Є., Карпінець В. В. «Комп‘ютерна стеганографія: навчальний посібник». Вінниця: ВНТУ. 2017.

2. Кузнецов О. О., Євсеєв С. П., Король О. Г. «Стеганографія: навчальний посібник». *Харків: Вид. ХНЕУ*. 2011.

3. Tyagi L. & Gupta A., Mohamed A. “Unveiling the invisible an in-depth analysis of text steganography techniques, challenges and advancement”. *3rd International Conference on Technological Advancements in Computational Sciences (ICTACS)*. Tashkent, Uzbekistan. 2023. p. 177–183. DOI: https://doi.org/10.1109/ICTACS59847.2023.10390024.

4. Ghosh S., Mondal R., Das A., Naskar A., Mullick S., Chatterjee S. “A study and reconsideration on text- steganography. international journal of research and analysis”. In *Science and Engineering (IJRASE*). 2023; 3 (6): 51–63.

5. Huang Y.-S., Tian C., Narayanan K., Zheng L. “Relatively-Secure LLM-Based steganography via constrained Markov decision processes”. *arXiv*. 2025. DOI: https://doi.org/10.48550/arXiv.2502.01827.

6. “Social Watermarking Ebook: The ultimate guide to social watermarking”. *Locklizard*. URL: https://www.locklizard.com/document-security-blog/ebook-social-watermarking (date of application: 24.09.2025).

7. Конахович Г. Ф., Прогонов Д. О., Пузиренко О. Ю. «Комп‘ютерна стеганографічна обробка й аналіз мультимедійних даних». *Підручник. Київ: «Центр учбової літератури»*. 2018.

8. “Microsoft. Visual Studio Code”. *Visual Studio Code*. URL: https://code.visualstudio.com (date of application: 24.09.2025).

9. “Python Releases for Windows”. *Python.org*. URL: https://www.python.org (date of application: 24.09.2025).

10. “Subprocess – Subprocess management”. *Python documentation*. URL: https://docs.python.org/uk/3/library/subprocess.html (date of application: 24.09.2025).

11. “PyPI”. *PyDes*. URL: https://pypi.org/pyDes (date of application: 24.09.2025).

12. “Hashlib - Secure hashes and message digests”. *Python documentation*. URL: https://docs.python.org/3/library/hashlib.html (date of application: 24.09.2025).

DOI: https://doi.org/10.15276/ict.02.2025.10
UDC 004.056.5:003.27

Implementation and comparative analysis of text steganography methods

Roman M. Khomiak[1)]
Master's Student of the Department of Information Technology Security
ORCID: https://orcid.org/0009-0007-0837-8402; roman.khomiak.mkbbi.2024@lpnu.ua
Viktor A. Melnyk[1)]
Doctor of Engineering Sciences, Professor of the Department of Information Technology Security
ORCID: https://orcid.org/0000-0002-5687-4204; viktor.a.melnyk@lpnu.ua
[1)] Lviv Polytechnic National University, 12 Stepan Bandera Str. Lviv. 79000, Ukraine

ABSTRACT

Steganography is a crucial area of modern information security, designed to conceal data within a carrier so that the very fact of secret transmission remains unnoticed. Among its various forms, text steganography holds a unique position due to the widespread use of the text format, though it faces a significant limitation: low container capacity compared to media like audio or images. Practically, text steganography is key to intellectual property protection, functioning as an "invisible social DRM" (Digital Rights Management). A unique, imperceptible identifier (marker) can be embedded into every copy of an e-book, which doesn't alter the content for the reader but allows for effective tracking of leakage sources or unauthorized distribution.

Considering the objective volume constraints of text as a container for hidden information, this article focuses specifically on the comparison and evaluation of methods. The goal was to select approaches that maximize the efficiency of data concealment while preserving the document's original structure and visual appearance. Our comprehensive study chose and practically implemented two core methods:

Homoglyph Substitution Method (using different variants of the same character that are visually indistinguishable to the reader but encode hidden data). This approach leverages the variability in character encoding standards (like Unicode) to embed information directly into the character code space, ensuring robustness against simple text copying.

End-of-Line Marker Modification Method (where variations in the formatting of paragraphs or lines subtly carry encoded information). This method exploits the redundancy in text file formats (e.g., ASCII/Unicode) regarding the indication of a line break (CR and LF).

The results of our comparative analysis convincingly demonstrate that text steganography, even with its inherent capacity limitations, is a promising and reliable tool for the individualization and protection of electronic publications. The application of these two methods ensures a high degree of imperceptibility for the hidden marker and achieves resilience against simple manipulation and copying of the content.

This combination of characteristics makes it possible to effectively scale DRM solutions (Social DRM), fostering new, secure approaches to the distribution of sensitive intellectual content. To practically showcase these advantages, a demonstration software implementation was created. This program, which utilizes both steganographic methods combined with additional encryption, clearly validates the potential for merging imperceptibility, resilience, and effectiveness.

Keywords: text steganography; text container; content protection; homoglyph substitution; line-ending character modification.