DESIGN OF INFORMATION PROTECTION SYSTEM USING LINGUISTIC STEGANOGRAPHY METHODS

Ovcharuk I.
PhD, Docent, Associate Professor of the Department of Information technologies, State University of Infrastructure and Technologies, Kyiv, Ukraine
Hoteev M.
Master's degree student State University of Infrastructure and Technologies, Kyiv, Ukraine

ПРОЄКТУВАННЯ СИСТЕМИ ЗАХИСТУ ІНФОРМАЦІЇ З ВИКОРИСТАННЯМ МЕТОДІВ ЛІНГВІСТИЧНОЇ СТЕГАНОГРАФІЇ

Овчарук І.
к.т.н., доцент, доцент кафедри інформаційних технологій Державного університету інфраструктури та технологій м. Київ, Україна
Хотєєв М.
магістрант, Державного університету інфраструктури та технологій м. Київ, Україна
https://doi.org/10.5281/zenodo.14008154

Abstract
The article examines methods of linguistic steganography that allow hiding text messages in Ukrainian-language texts by alternating prepositions and conjunctions in accordance with the rules of Ukrainian spelling. Linguistic steganography provides covert transmission of data without an obvious signal of its existence, which reduces the probability of detection of a secret message.

Анотація
У статті досліджено методи лінгвістичної стеганографії, які дозволяють приховувати текстові повідомлення в україномовних текстах через чергування прийменників і сполучників відповідно до правил українського правопису. Лінгвістична стеганографія забезпечує приховану передачу даних без явного сигналу про їх існування, що зменшує ймовірність виявлення секретного повідомлення.

Keywords: linguistic steganography, information hiding, information security, encoding, decoding, steganoanalysis, digital security, text data
Ключові слова: лінгвістична стеганографія, приховування інформації, інформаційна безпека, кодування, декодування, стеганоаналіз, цифрова безпека, текстові дані

Вступ. У сучасному світі ми постійно взаємодіємо з величезними обсягами інформації. Не дарма кажуть: «Хто володіє інформацією, той володіє світом!». Люди роблять нові відкриття, розвивають технології, спілкуються в чатах, створюють культурні продукти, такі як книги, пісні, фільми. Відтепер ділитися своїми досягненнями стає зручніше завдяки мережевим технологіям. Однак, багато хто не задумується про те, що їхнє приватне життя може стати об'єктом уваги зловмисників. Зазвичай люди думають: «Я не політик, не олігарх, не вчений. Мене це не стосується!». Проте це хибне уявлення, що легко спростовується новинами або статтями в Інтернеті. Існує багато загроз: шахрайство з використанням особистих даних, промисловий шпіонаж, порушення авторських прав, викрадення державних секретів або навіть шантаж. Оскільки велика частина інформації зберігається та передається у вигляді текстів, не дивно, що людство здавна приховувало повідомлення саме в них.

Стеганографія – це наука, що займається приховуванням інформації, а лінгвістична стеганографія спеціалізується на приховуванні даних у текстах [1]. Вона використовує звичайні тексти як «контейнери», які не повинні викликати підозри. Суть лінгвістичної стеганографії полягає у тому, щоб закодувати будь-яку інформацію у будь-якому тексті, спираючись на лінгвістичні засоби. Однією з переваг цього методу є те, що секретне повідомлення можна передавати як під час розмови, так і через лист або рукописний текст [2].

Аналіз останніх досліджень і постановка проблеми. Останнім часом спостерігається значний інтерес до методів захисту інформації в умовах швидкого розвитку цифрових технологій та збільшення обсягів переданих даних [3, 4]. Традиційні методи шифрування та криптографії залишаються ключовими інструментами в інформаційній безпеці, проте вони часто привертають увагу зловмисників, оскільки наявність зашифрованого повідомлення сама по собі сигналізує про важливість даних. У зв'язку з цим усе більшого значення набуває стеганографія, яка дозволяє приховати факт передачі секретної інформації [5, 6].

Одним із перспективних напрямків у цій галузі є лінгвістична стеганографія, яка використовує природні мови для приховування даних у текстах [7, 8]. Вона дозволяє вбудовувати приховані повідомлення в звичайні тексти таким чином, що ці тексти не викликають підозр у сторонніх осіб. Ряд дослідників, таких як М. Фрідман, Г. Блум, та інші, досліджували різні підходи до приховування інформації у текстах, зокрема за допомогою семантичних і синтаксичних змін. Проте, попри прогрес у цій

сфері, лінгвістична стеганографія все ще залишається складним напрямком для реалізації на практиці через високі вимоги до збереження природності тексту та забезпечення стійкості до атак на виявлення прихованої інформації.

Основною проблемою є розробка таких методів лінгвістичної стеганографії, які б дозволяли не тільки приховувати повідомлення з високим рівнем секретності, але й мінімізувати ризики їх виявлення за допомогою сучасних алгоритмів стеганоаналізу [9, 10]. Крім того, необхідно створювати системи, здатні працювати з великими обсягами текстових даних, забезпечуючи водночас високу точність приховування та захист від спроб витоку інформації.

Мета і завдання дослідження. Метою даного дослідження є проєктування системи для кодування текстових повідомлень та їх приховування в інших текстових файлах за допомогою методів лінгвістичної стеганографії. Особливістю підходу є врахування специфіки української мови, зокрема правил написання та чергування прийменників і сполучників. Це дозволить створювати стеганографічні повідомлення, які природно вписуються в україномовні тексти, зберігаючи високу прихованість інформації.

Об'єктом дослідження є процес стеганографічного приховування інформації в текстах українською мовою, а предметом – алгоритм приховування текстової інформації, який ґрунтується на чергуванні прийменників і сполучників відповідно до правил українського правопису.

Матеріали та методи дослідження. Алгоритм стеганографічного приховування інформації в тексті ґрунтується на двох основних підпрограмах: кодування та декодування. Кожна підпрограма має свою функціональну структуру, розділену на блоки, що забезпечують кодування текстового повідомлення та його приховування в іншому тексті. Розглянемо ці процеси детальніше з використанням формул та математичних моделей для більшої точності опису.

Кодування полягає у перетворенні текстового повідомлення у двійковий формат і подальшому його інтегруванні в текст через заміну певних пар символів. Алгоритм кодування наведено на рисунку 1.

Робота з повідомленням:

1. Перетворення символів у числові значення. Для кожного символу повідомлення, програма виконує його перетворення в ASCII-код:

$$ASCII(M_i) = int(M_i), \quad (1)$$

де M_i – символ повідомлення, а $ASCII(M_i)$ – його числовий код.

2. Перетворення у двійкову систему. Отримані ASCII-коди перетворюються у двійковий формат:

$$B_i = bin(ASCII(M_i)), \quad (2)$$

де B_i – двійкове представлення символу.

Наприклад, якщо символ «А» має ASCII-код 65, його двійкове представлення буде:

$$B_{\text{А}} = 01000001_2$$

Робота з текстом:

1. Обробка текстового файлу. Програма циклічно опрацьовує текстовий файл, знаходячи пари символів для заміни. Цей процес триває до завершення файлу:

$$T = \sum_{i=1}^{N} t_i, \quad (3)$$

де T – текст, що обробляється, а N – кількість символів у файлі.

2. Знаходження груп символів. Програма ідентифікує такі пари символів: «у» та «в», «і» та «й». Кожна пара може використовуватися для передачі одного біта даних:

- якщо біт дорівнює «0», програма замінює пару на «у» або «і»;
- якщо біт дорівнює «1», пара замінюється на «в» або «й».

Формально це можна записати як функцію заміни:

$$S(b) = \begin{cases} \{\text{у}, \text{і}\}, \text{якщо } b = 0 \\ \{\text{в}, \text{й}\}, \text{ якщо } b = 1 \end{cases}, \quad (4)$$

де b – біт, що передається.

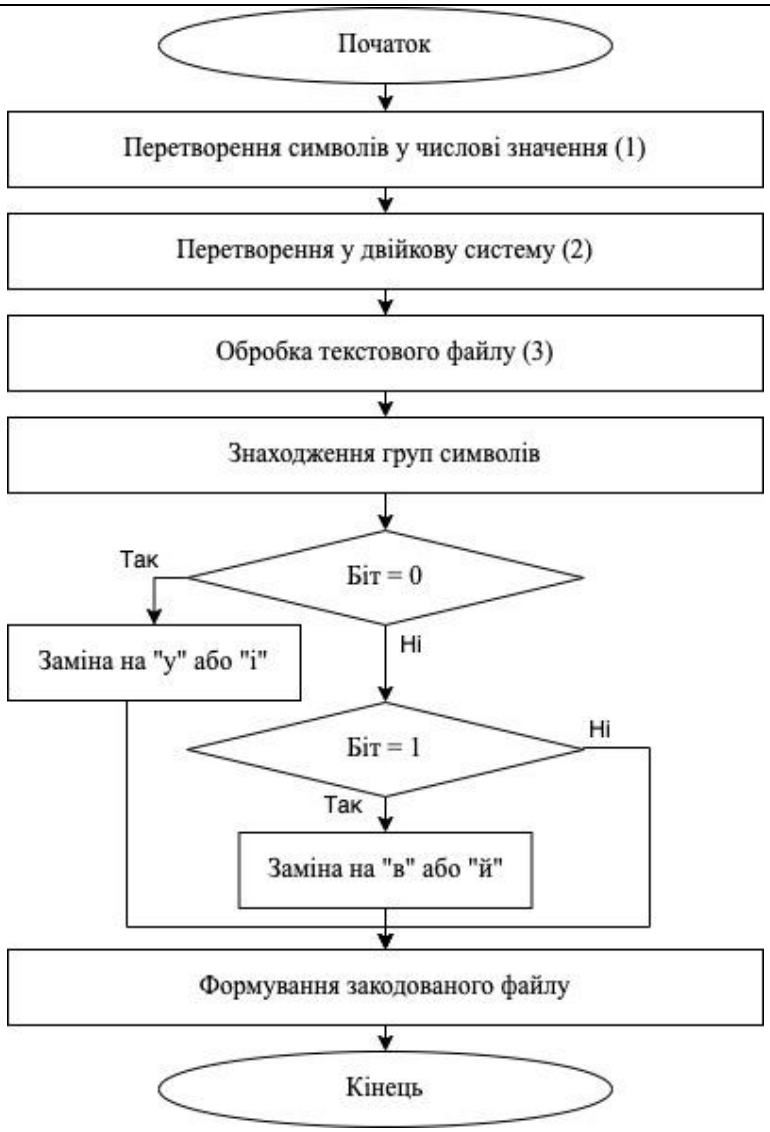

Рис 1. Алгоритм кодування

3. Формування закодованого файлу. У результаті обробки тексту з урахуванням бітів двійкової системи вихідного повідомлення, програма генерує закодований файл. Всі проміжні файли видаляються після завершення роботи.

Даний алгоритм кодування надає чітке розуміння процесу кодування, від перетворення символів у числові значення до формування закодованого файлу.

Процес декодування полягає у зворотному перетворенні двійкових даних у текстове повідомлення. Алгоритм декодування наведено на рисунку 2.

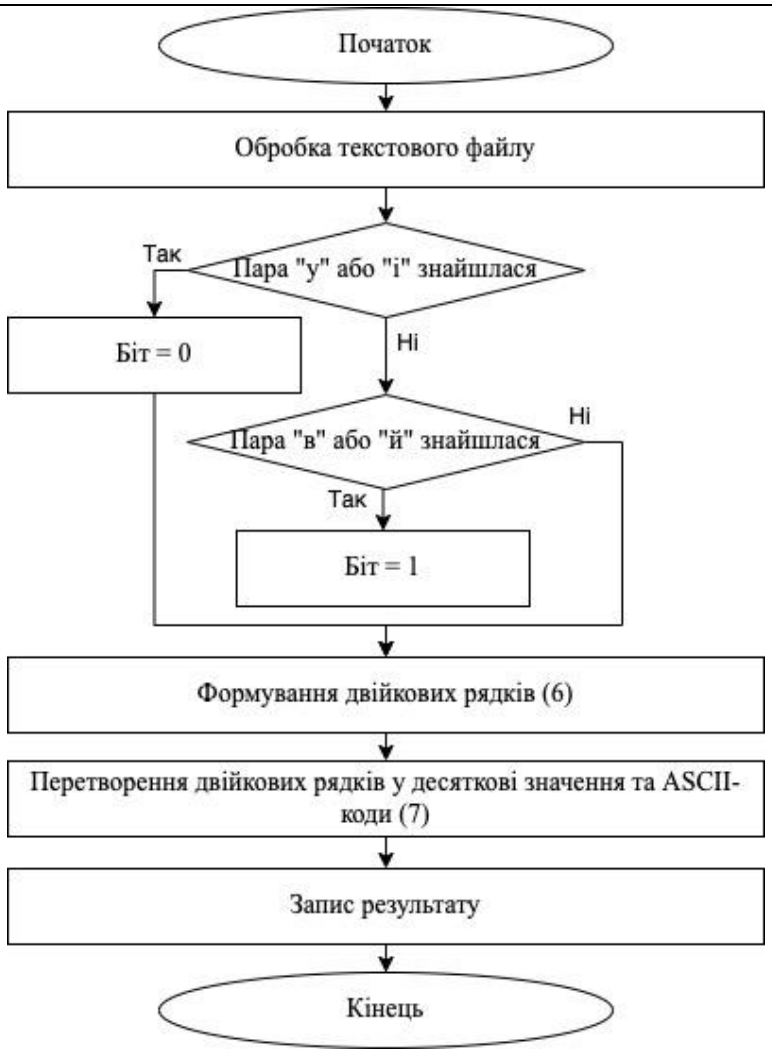

Рис 2. Алгоритм декодування

1. Обробка текстового файлу. Як і в процесі кодування, програма циклічно аналізує текстовий файл, знаходячи пари символів: «у» та «в», «і» та «й».

2. Ідентифікація двійкових бітів. Якщо програма знаходить пару «у» або «і», то це відповідає біту «0». Якщо знаходяться пари «в» або «й», це означає біт «1»:

$$b = \begin{cases} 0, \text{якщо } \{\text{у, і}\} \\ 1, \text{ якщо } \{\text{в, й}\}' \end{cases} \quad (5)$$

3. Формування двійкових рядків. Програма формує рядки двійкових чисел довжиною 8 бітів, що відповідають одному символу:

$$M_i = B_1 B_2 \dots B_8, \quad (6)$$

де M_i – символ повідомлення, а $B_1 B_2 \dots B_8$ – відповідні біти.

4. Перетворення з двійкової в десяткову систему та ASCII-коди. Двійкові значення перетворюються у десяткову систему і далі – у текстові символи:

$$ASCII^{-1}(B_i) = M_i, \quad (7)$$

де $ASCII^{-1}$ – зворотне перетворення з ASCII-коду у символ.

5. Запис результату. Після завершення процесу декодування програма формує вихідний текстовий файл з відновленим повідомленням. Всі проміжні файли видаляються.

Даний алгоритм декодування дозволяє відновити початковий текст, який було закодовано, шляхом заміни символів на пари «у», «в», «і» та «й».

Загальна схема роботи алгоритму може бути подана через функції кодування та декодування:

1. Кодування:

$$C(T, M) = \dot{T}, \quad (8)$$

де C – функція кодування, T – текстовий контейнер, M – повідомлення, що кодується, $\dot{T}$ – закодований текст.

2. Декодування:

$$D(\dot{T}) = M, \quad (9)$$

де D – функція декодування, $\dot{T}$ – закодований текст, M – відновлене повідомлення.

Таким чином, розроблений алгоритм забезпечує ефективне приховування інформації в тексті за допомогою лінгвістичної стеганографії, використовуючи особливості чергування прийменників та сполучників в українській мові.

Висновки. У ході дослідження було розроблено та проаналізовано алгоритм лінгвістичної стеганографії, що дозволяє приховувати текстові повідомлення в україномовних текстах за допомогою чергування прийменників і сполучників відповідно до правил українського правопису. Запропонована система кодування та декодування інформації забезпечує високу ефективність приховування даних

та підтримує природність тексту, що дозволяє мінімізувати ризик виявлення прихованої інформації сторонніми особами.

Особливість розробленого алгоритму полягає в його здатності використовувати особливості української мови для стеганографії, що робить його адаптованим до специфіки мови й відповідним для застосування у реальних умовах. При цьому досягнуто балансу між складністю реалізації та надійністю приховування повідомлень. Важливим аспектом є також можливість застосування алгоритму до великих обсягів текстових даних.

Отримані результати можуть слугувати основою для подальших досліджень у напрямку лінгвістичної стеганографії, зокрема щодо розширення можливостей алгоритму для інших мовних груп та удосконалення методів захисту від стеганоаналізу.

Список літератури:

1. Кузнецов О.О. Стеганографія: навчальний посібник / О.О. Кузнецов, С.П. Євсеєв, О.Г. Король. Х. : Вид. ХНЕУ, 2015. 232 с.

2. Хорошко В.О. Комп'ютерна стеганографія: навчальний посібник / В.О. Хорошко, Ю.Є. Яремчук, В.В. Карпінець. Вінниця : ВНТУ, 2017. 155 с.

3. Козюра В.Д. Захист інформації в комп'ютерних системах :підручник / В.Д. Козюра, В.О. Хорошко, М.Є. Шелест. Ніжин : ФОП Лукяненко В.В., ТПК «Орхідея», 2020. 236 с.

4. Конахович Г.Ф. Компютерна стеганографічна обробка й аналіз мультимедійних даних : підручник / Г.Ф. Конафович, Д.О. Прогонов, О.Ю. Пузиренко. К. «Alex Print Centre», 2018. 558 с.

5. Бурячок В.Л. Інфомаційна та кібербезпека / В.Л. Бурячок, В.Б. Толубко, В.О. Хорошко, С.В. Толюпа. К.: ДУТ, 2015. 288 с.

6. Мельник С. Методи цифрової стеганографії: стан та напрями розвитку // С. Мельник, В. Кащук. // Information Security of the Person, Society and State. 2013. No3. С. 65–70.

7. Vovk, O. Synthesis of optimal steganographic method meeting given criteria / O. Vovk, A. Astrahantsev // Informatyka Automatyka Pomiary w Gospodarce i Ochronie Środowiska (technical and scientific journal). Lublin, Poland, 2015. P. 27–34

8. Kovtun, Vladislav & Gnatyuk, Sergiy & Кінзерявий, Олексій. (2013). Modern methods of computer steganography systematization. Ukrainian Scientific Journal of Information Security. 19. DOI: 10.18372/2225-5036.19.5628.

9. Нога О.В. Методи стеганографічного захисту і стеганоаналізу інформації з використанням аудіофайлів: кваліфікаційна робота бакалавра за спеціальністю 125 – Кібербезпека / О.В. Нога. Тернопіль : ТНТУ, 2022. 52 с.

10. Xiao Steganography. [Електронний ресурс]. Режим доступа: http://download.cnet.com/Xiao-Steganography/3000-2092_4-10541494.html