© Fedotova-Piven I., Tarasenko Y The ways of relevant of modern cybersecurity in context of countering the methods of computer linguistic steganography // Ukrainian Scientific Journal of Information Security, 2017, vol. 23, issue 3, p.190-196.

# СТЕГАНОГРАФІЯ ТА СТЕГОАНАЛІЗ / STEGANOGRAPHY & STEGANALYSIS

**DOI:** 10.18372/2225-5036.23.12093

# ШЛЯХИ ЗАДОВОЛЕННЯ ПОТРЕБ СУЧАСНОЇ КІБЕРБЕЗПЕКИ В РАМКАХ ПРОТИДІЇ МЕТОДАМ КОМП'ЮТЕРНОЇ ЛІНГВІСТИЧНОЇ СТЕГАНОГРАФІЇ

**Ірина Федотова-Півень, Ярослав Тарасенко**

*Черкаський державний технологічний університет*

**ФЕДОТОВА-ПІВЕНЬ Ірина Миколаївна,** к.т.н.

*Дата та місце народження:* 1973 рік, м. Сміла, Черкаська обл., Україна.
*Освіта:* Черкаський державний технологічний університет, 2004 рік.
*Посада:* завідувач кафедри інформаційної безпеки та комп'ютерної інженерії.
*Наукові інтереси:* багатооперандне додавання в рекурентних системах числення.
*Публікації:* більше 30 друкованих праць.
*E-mail:* irynapiven73@gmail.com

**ТАРАСЕНКО Ярослав Володимирович**

*Рік та місце народження:* 1993 рік, м. Черкаси, Україна.
*Освіта:* Черкаський державний технологічний університет, 2015 рік; Черкаський державний технологічний університет, 2016 рік
*Посада:* аспірант кафедри інформаційної безпеки і комп'ютерної інженерії.
*Наукові інтереси:* комп'ютерна лінгвістична стеганографія, математична та прикладна лінгвістика, інформаційна безпека, комп'ютерні системи.
*Публікації:* 6 наукових публікацій.
*E-mail:* yaroslav.tarasenko93@gmail.com

***Анотація.*** *В статті проводиться огляд поширених методів текстової стеганографії, а саме методів довільного інтервалу, синтаксичних і семантичних методів, досліджуються існуючі шляхи протидії їм, а також засоби автоматизованого лінгвістичного аналізу тексту (морфологічного, синтаксичного, дискурсного) для автоматизації текстового стегоаналізу. Проводиться узагальнення існуючих методів та відзначаються такі, які можуть набути подальшого розвитку та вдосконалення при вирішенні задачі протидії засобам комп'ютерної лінгвістичної стеганографії, і можуть бути реалізовані за допомогою лінгвістичного процесору (метод логічного множення для морфологічного, імовірнісно-статистичний метод для синтаксичного та текстуально-інтертекстуальний підхід для дискурсного аналізу). Таким чином доводиться необхідність проведення наукового дослідження і створення методу стиснення текстової інформації для лінгвістичної стеганографії.*

***Ключові слова****: кібербезпека, лінгвістична стеганографія, методи довільного інтервалу, синтаксичні методи, семантичні методи, морфологічний аналіз, синтаксичний аналіз, дискурсивний аналіз, лінгвістичний процесор, автоматизований стегоаналіз.*

## Вступ

Ера інформаційних технологій сприяє глобалізації суспільства, руйнує бар'єри в передачі інформації, але крім користі існує безліч прихованих загроз. Будь-яка текстова інформація, що передається в мережі Інтернет може містити в собі приховане повідомлення, а об'єм текстової інформації, що передається щосекунди унеможливлює ефективну перевірку наявності стегоповідомлення в ній. Стеганографія вносить значні корективи в уявлення успішної передачі повідомлення порівняно з криптографією, оскільки приховане повідомлення не лише повинно бути не виявлене, але навіть про його існування ніхто не повинен здогадатися [1, с.9].

Кількість та якість методів приховування повідомлення в тексті щоденно зростає. Текстова стеганографія може мати різноманітний вигляд: від зміни

форматування вже існуючого тексту, модифікації пунктуації до заміни слів у тексті та генерації тексту програмними стеганографічними засобами [1, с.7].

Це зумовлює широкий спектр потреб кібербезпеки в рамках протидії методам комп'ютерної лінгвістичної стеганографії, що потребують вирішення. А подальший розвиток шляхів задоволення цих потреб може забезпечити ефективну протидію витокам та незаконній міграції секретної інформації в результаті використання методів текстової стеганографії.

Такий широкий спектр можливостей для атак ускладнює імовірність виявлення прихованого повідомлення та зумовлює необхідність створення комплексних програмних засобів стегоаналізу та засобів знищення стегоповідомлення, навіть за умови, що напевно невідомо чи присутнє воно у тексті.

Саме те, що наявність прихованого повідомлення у тексті не повинна бути розкрита, підштовхує до пошуку методів виявлення самого факту наявності такого повідомлення, шляхів його розшифрування, та/або спотворення чи повного його знищення, інакше кажучи, сприяє розвитку стегоаналізу.

Оскільки, автоматизований аналіз тексту вимагає використання лінгвістичних методів, це викликає потребу звертатися також до методів комп'ютерної лінгвістики, а саме до автоматизованого аналізу тексту від морфологічного аналізу конкретного слова до дискурс аналізу тексту чи навіть сукупності текстів, а також до суміжних дисциплін.

**Аналіз досліджень та постановка завдання**

У роботах сучасних дослідників в області комп'ютерної лінгвістичної стеганографії постійно робляться спроби удосконалення існуючих методів стегоаналізу тексту природньої мови, як у працях Рябко Б.Я. [2], які удосконалив Нечта І.В. [3] та проводяться дослідження в напрямку розробки новітніх альтернативних засобів, як у роботі Рошіді Дін та ін. [4], що пропонує словниковий підхід. Проте, нажаль, існує досить мало методів стегоаналізу текстів природньої мови [4].

Базуючись на останніх дослідженнях, можна виділити невирішені частини загальної проблеми. Так, Нечта І. В. в статті «Ефективний метод стегоаналізу, що базується на стисненні даних» [3] звертає увагу на недолік методів, що генерують текст, а саме отримання неосмисленого тексту [3, с.51]. Такий неосмислений текст буде відповідати усім правилам граматики та не викликатиме жодних підозр в автоматизованих систем стегоаналізу. Автор наголошує на актуальності створення ефективних засобів автоматизованого комп'ютерного аналізу [3, с.51], а також пропонує метод виявлення наявності стегоповідомлення шляхом стиснення інформації звичайним архіватором, суть якого виявити стегоповідомлення, базуючись на порушенні статистичної структури контейнеру [3, с.53]. Недоліком запропонованого підходу є його неефективність, коли йдеться про текстове повідомлення, згенероване автоматично за допомогою програмних засобів генерації тексту. В такому випадку стиснення файлу не дасть ніяких результатів, адже контейнер зі стегоповідомленням буде вбудований шляхом маніпуляцій на рівні смислової інформації, хоча текст буде і неосмислений, але для систем аналізу він не викликатиме підозр. В такому разі переваги методу, описані автором, а саме порівняно невеликий час аналізу та відсутність необхідності використання великих за об'ємом займаної пам'яті словників [3] зводяться до нуля. Проте описана методика може бути використана в якості основи для смислового стиснення текстової інформації, що може ефективно діяти в умовах зростаючого об'єму передачі даних мережею Інтернет.

В свою чергу проводиться робота по протидії атаці стисненням, що описував Нечта І.В., таким чином, Мельник М.А. в дисертації на тему «Підвищення стійкості стеганографічної системи до атаки стисненням» [5] проводить роботу над побудовою теоретичного базису і розробкою на його основі методів і алгоритмів, стійких до стиснення зі значними коефіцієнтами для підвищення ефективності роботи стегосистеми [5], проте розробки в області смислового стиснення інформації ведуться в недостатньому обсязі і атака стисненням саме смислової інформації матиме значний успіх в умовах відсутності засобів протидії подібним атакам.

Більш детальний огляд основних методів текстової стеганографії, а також способів боротьби з ними зможе дати чітке уявлення про подальші перспективи розробки стегоаналітичних методів та їх реалізації у вигляді програмних комплексів і застосування їх на практиці для виявлення, спотворення, модифікації чи знищення стегоповідомлення, а відповідно задоволення потреб кібербезпеки.

Актуальність роботи обумовлена тим фактом, що однією з важливих потреб кібербезпеки є розробка ефективних методів стегоаналізу текстової інформації, адже, існує досить мало методів стегоаналізу текстів природньої мови [4, с.444-445]. Це зумовлено зростаючою потребою в забезпеченні боротьби з витоком секретної інформації із комерційних чи державних структур, а враховуючи об'єм інформації, що передається, перспективними є саме автоматизовані методи аналізу тексту [4, с.245]. В той же час, для визначення шляхів задоволення потреб кібербезбеки та виділення майбутніх перспектив виникає необхідність в узагальненні методів автоматизованого лінгвістичного аналізу тексту, а також визначенні найбільш доцільних методів в рамках протидії засобам комп'ютерної лінгвістичної стеганографії.

Метою дослідження є огляд поширених методів текстової стеганографії та шляхів боротьби з ними. Виділення найбільш відповідних засобів протидії методам комп'ютерної лінгвістичної стеганографії серед існуючих методів автоматизованого аналізу тексту для визначення шляхів задоволення потреб сучасної кібербезпеки у виявленні та нейтралізації загроз, зумовлених впровадженням методів комп'ютерної лінгвістичної стеганографії.

**Основна частина дослідження**

Існує багато способів передачі прихованого повідомлення, деякі методи передбачають модифікацію вже існуючого тексту, інші генерують власний текст, проте всі вони мають свої недоліки, а ефективність передачі повідомлення та об'єм даних, що

можливо передати прямо пропорційно залежить від наявності методів стегоаналізу [6 с.245].

**Найбільш поширені методи текстової стеганографії та шляхи боротьби з ними.** Говорячи про текстову стеганографію не можна не згадати методи описані в статті Крижановської О.Л. «Аналіз методів текстової стеганографії» [7]. Автор виділяє 3 основні групи: методи довільного інтервалу (метод зміни кількості пропусків між реченнями, метод зміни кількості пропусків у кінці текстових рядків, метод зміни кількості пропусків між словами вирівняного за шириною тексту), синтаксичні і семантичні методи [7]. Як вже зрозуміло, методи, що відносяться до першої групи мають досить низьку криптостійкість особливо коли йдеться про автоматизований аналіз, адже зайві інтервали легко помітити і виділити закономірності їх розподілу. Синтаксичні методи полягають в зміні пунктуації, абревіатури і скорочень тексту [7], що можна вважати більш криптостійкими особливо до засобів автоматизованого стегоаналізу, адже в такому випадку слід проводити дослідження в напрямку смислової цілісності тексту. Семантичні методи є найбільш стійкими. Застосовуються різні підходи маніпулювання самим реченням і словами [7], зокрема автор відносить до цього напрямку метод заміни синонімів. Протидіяти таким методам значно важче, текст слід аналізувати також на рівні смислу та розглядати текст в цілому, що значно важче реалізувати на програмному рівні.

Іванов І.Г. в статті «Використання методів текстової стеганографії» [8] також описує вищезгадані методи, але вносить деякі доповнення. Зокрема, автор виділяє ще одну групу стеганографічних методів, а саме мімікрію, що полягає у генерації штучного осмисленого тексту і вбудовуванні в нього інформації шляхом відбору певних фраз і слів [8, с.173]. Таким чином, згенерований текст не матиме ніяких явних відхилень ні в синтаксичній, ні в морфологічній структурі лексичних одиниць, ні в форматуванні, пунктуація також не викликатиме ніяких підозр. В такому разі слід аналізувати саме дискурс з метою виявлення відсутності завершеності тексту, наявності розбіжностей між його мікротемами та безпосередньо виявляти мету його написання.

В свою чергу Кошева Н.А. в монографії «Інформаційні технології і захист інформації в інформаційно-комунікаційних системах» [6] під час аналізу мімікрії вносить деякі уточнення, які стосуються зокрема популярних програм, що генерують штучний текст – Nicetext, Texto и Markov-Chain-Based [6, с.244] та зазначає, що результатом роботи цих програм є беззмістовний текст. Також автор стверджує, що аналіз осмисленості тексту можна проводити лише за участі людини, що не завжди можливо через великий об'єм інформації [6, с.244]. Тим самим підтверджує актуальність та необхідність розробки автоматизованих засобів смислового аналізу великих об'ємів текстів.

Тож можна вважати ці методи класичними, проте стеганографія стрімко розвивається і постійно з'являються інноваційні методи. Пошук шляхів протидіяти їм є головним завданням, оскільки відсутність ефективних засобів стегоаналізу гарантує дієвість таких методів.

Так, можна виділити деякі новітні напрями текстової стеганографії, як жаргонні шифри, що представляють особливий інтерес, оскільки тут використовуються мовні засоби приховування інформації [9, с.68], ефективно протидіяти яким можливо тільки шляхом визначення смислу повідомлення.

Існує цілий ряд методів стегоаналізу, що в тій чи іншій мірі здатні ефективно протидіяти методам стеганографії, проте лише частину з них можливо використовувати для задоволення потреб саме текстового стегоаналізу. Виділяють такі групи методів, що розділяються по об'єкту пошуку в стегоконтейнері: методи сигнатурного типу (візуальна атака на стегосистему, аналіз відповідності формату даних), імовірнісні методи (аналіз статистики Хі-квадрат, метод дослідження статистичних моделей вищого розряду); по впливу на файл контейнер: методи пасивного і активного стегоаналізу; по широті аналізу контейнерів: методи, призначені для виявлення даних, що приховані за допомогою конкретного алгоритму, методи «сліпого» розпізнання [10].

Таким чином, Нечта І.В. в статті «Використання статистичного аналізу для виявлення прихованих повідомлень в текстових даних» [11] демонструє працездатність та ефективність методу стегоаналізу Хі-квадрат для дослідження текстових даних. Цим самим автор доводить універсальність методів стегоаналізу, проте, як відомо, універсальні системи не завжди можуть бути ефективними у конкретних випадках. Ця проблема вирішується комплексним підходом.

Проте, відомо що існує 2 роди помилок програм стегоаналізу: прийняття пустого контейнеру за заповнений та прийняття заповненого контейнеру за пустий [12, с.30]. З використанням комплексного підходу до стегоаналізу можна імовірність помилки будь-якого роду значно зменшити, проте неможливо досягти повної відсутності цих помилок. Зменшення імовірності помилки можливо досягти за допомогою методів, що не просто протидіють відомим засобам стеганографії, а стоять, так би мовити над усією системою в комплексі та можуть протидіяти так само ефективно і невідомим методам приховування інформації, що можуть з'явитися в майбутньому.

Отже, кібербезпека потребує проведення досліджень в напрямку розробки сугубо лінгвістичних методів стегоаналізу, зокрема що відноситься до сфери дослідження смислового навантаження текстового повідомлення, а також не мало важливою залишається проблема інтеграції їх з діючими математичними алгоритмами стегоаналізу.

**Лінгвістичний процесор та його структура.** Потреба кібербезпеки в автоматизованому стегоаналізі полягає в оцінці осмисленості текстового повідомлення. Це досягається в процесі використання так званого лінгвістичного процесора. Андрєєв А.М. в статті «Імовірнісний синтаксичний аналізатор для інформаційно-пошукової системи» [13] визначає таку задачу лінгвістичного процесору: це «перетворення речення природньою мовою (або навіть цілого тексту) в певний набір семантичних структур, що

є формальним поданням «сенсу» вихідного речення або тексту» [13]. Автор описує класичну структуру лінгвістичного процесору, що складається з трьох послідовних блоків для морфологічного, синтаксичного та семантичного аналізу та стверджує, що може існувати також блок лексичного аналізу, що здійснює фрагментацію тексту на речення, а потім на слова, числа і знаки пунктуації [13].

Хоча описаний лінгвістичний процесор використовується для роботи інформаційно-пошукових систем, його структуру можна взяти за основу та адаптувати під задачі та потреби стегоаналізу. Основною відмінністю буде заміна третього блоку класичної моделі, а саме семантичного аналізатору на модуль дискурс-аналізу тексту.

Описаний підхід може бути взятий за основу для виділення суті текстового повідомлення, подальшого відсіювання другорядної інформації, виокремлення головної думки тексту. Блок так званого лексичного аналізу можна використати з метою виявлення закономірностей розподілу знаків пунктуації, порівняння з існуючими правилами граматики. На основі отриманих даних виявити повну картину використання механічних засобів стеганографії в досліджуваному тексті.

З іншого боку, задовольнити потребу кібербезпеки в аналізі осмисленості тексту, виділенні його основної теми та аналізі його структури чи послідовності, визначенні наявності слідів модифікації стеганографічними засобами можливо шляхом побудови лінгвістичного процесору, що складатиметься з морфологічного синтаксичного та дискурсного аналізаторів, які в комплексі зможуть оцінити осмисленість інформації, а окремо ефективно протидіяти різноманітним методам комп'ютерної лінгвістичної стеганографії. Саме дискурс аналіз, на відміну від інших методів дослідження тексту розглядає текст як єдине ціле та дозволяє отримати також мету написання тексту, адже, як стверджує Прокошенкова Л.П. в статті «Дискурсивний аналіз та його роль в сучасній лінгвістиці» [14], метою дискурсного аналізу виступає виокремлення соціального аспекту [14, с.452] написання того чи іншого тексту, інакше кажучи, такий блок текстового процесору дозволить виокремити саму мету написання тексту, що доводить необхідність його використання в області стегоаналізу, адже такою метою може виступати саме передача прихованого повідомлення.

Для підвищення ефективності стегоаналізу слід провести огляд найбільш поширених методів побудови кожного блоку лінгвістичного процесору, шляхи їх автоматизації та обґрунтувати вибір тих, що зможуть найповніше задовольни потреби стегоаналізу.

**Методи автоматизованого лінгвістичного аналізу тексту.** Для розробки ефективних алгоритмів стегоаналізу безпосередньо текстової інформації слід перш за все звернутися до лінгвістичних методів автоматизованого аналізу тексту. У відповідь на методи вбудовування інформації в текстове повідомлення необхідно розглянути відповідні шляхи протидії їм. Так, належить аналізувати підозрілий текст морфологічно, синтаксично та дискурсно. Проте існує не один метод кожного з наведених напрямів аналізу, а як було згадано, саме вибір оптимального методу зумовлює ефективність стегоаналізу.

Говорячи про морфологічний аналіз тексту належить згадати, що Бабіна Ольга Іванівна в статті «Корпусний метод автоматичного морфологічного аналізу флективних мов» [15] стверджує про наявність двох типів морфологічних аналізаторів, що основуються на словникових системах, а саме системи з базою лексем та системи з базою словоформ, а також безсловникові системи [15, с. 39]. Звісно, кожна із систем має свої переваги та недоліки, наприклад система, що використовує словникову базу буде прямо пропорційно сповільнювати роботу аналізатору за умови збільшення об'єму аналізованого тексту. Робота ж безсловникових систем навпаки вимагає великого об'єму тексту, оскільки алгоритм роботи основується на математичних алгоритмах машинного навчання [15, с.40]. Також в статті подається розробка універсального корпусного методу, який би на думку автора включав усі переваги та недоліки кожної з систем, а також зазначається трудомісткість методу, оскільки він вимагає ручного збору лінгвістичних даних, хоча відмінною особливістю є факт визначення лише флексій [15, с.40]. Веб-сайт «Лінвістика» [16] в свою чергу доповнює описані данні, а саме виділяється третій тип морфологічного аналізу зі словником – аналіз методом логічного множення. Проте, для вирішення конкретних задач стегоаналізу універсальність підходить не завжди, адже на ряду з ефективністю самого методу морфологічного аналізу не слід забувати про трудомісткість процесу та час виконання операцій.

Використання кожного з методів зумовлюється особливостями мови, якою написано аналізований текст. Наприклад, морфологічний аналіз зі словником словоформ використовується для мов з бідною морфологією, а зі словником основ для більшості європейських мов [16].

Оскільки, при використанні різноманітних методів аналізу текстової інформації в комплексі важливо не забувати про наступний етап, тому вибір методу морфологічного аналізу повинен узгоджуватись саме із можливістю ефективної взаємодії із синтаксичним аналізом речень, до того ж цільова мова для аналізу не завжди відноситься до європейської групи, а досить часто належить до слов'янсьої групи. На основі цього, та базуючись на тому факті, що виділення закінчень та основ слова може покращити визначення логічних зв'язків у реченні, то в конкретному випадку найкраще використовувати саме метод логічного множення. Він хоч і потребує використання словника словоформ, проте однаково ефективно підійде до будь-якої мови аналізу та дозволить покращити та спростити наступні етапи аналізу тесту.

Далі, на базі отриманих результатів, можна виявити зв'язки між словами, згідно з якими визначити головні і другорядні члени речення, тобто провести синтаксичний аналіз тексту та отримати ширшу інформацію про сам текст, щоб з більшою точністю зробити висновок про модифікацію його за допомогою стеганографічних методів.

Проте, існує вибір способів автоматичного синтаксичного аналізу, кожен з яких володіє своїми перевагами та недоліками, що критично впливають на процес стегоаналізу. Таким чином, Аношин П. І. в статті «Автоматичний аналіз текстів. Синтаксичний і семантичний аналіз» [17] описує наявність двох підходів вирішення проблеми неоднозначності синтаксису: формально-графічний та імовірнісно-статистичний [17]. Кожен з підходів зумовлює структуру синтаксичного аналізатору та його особливості. Автор визначає суть формально-графічного підходу, яка полягає в створенні складних правил, згідно з якими можна визначати граматичну структуру та імовірнісно-статистичного, де відбувається збір статистики вживання певних граматичних структур у схожих текстах [17]. Також, автор зазначає, що методи імовірнісного аналізу не здатні повністю забезпечити стовідсоткову точність аналізу, проте задовільні для роботи з реальними текстами, та згадує про перевагу імовірнісного аналізу в плані затрат на розробку [17]. Виходячи з цих характеристик, можна стверджувати, що безпосередньо для стегоаналізу найбільш ефективно буде використовувати саме імовірнісно-статистичний підхід до синтаксичного аналізу. По-перше, це пов'язано з тим, що мета синтаксичного аналізу полягає не в перекладі тексту, не вимагає штучного синтезу, а лише є частиною системи для визначення осмисленості тексту та виділення його змісту, а це не вимагає повної точності у процесі визначенні дерева відповідностей синтаксичних структур, і на ряду з цим здешевить розробку та знизить її складність. По-друге, формально графічний метод не є гнучким, що ускладнює його роботу з реальними текстами, особливо розмовною мовою.

Вищезгадані два підходи передбачають майбутню побудову та використання дерева відповідностей. Але, Войтех Ковар в дисертації «Автоматичний синтаксичний аналіз для реальних додатків» [18] доводить, що розмітка у вигляді звичайного чи XML тексту, що використовується для запису часткової синтаксичної інформації хоч і володіє меншою виразністю, ніж у випадку синтаксичних дерев, проте прозорість забезпечує зручність практичного використання методу [18, с.24], що немало важливо у випадку створення прикладного комплексу для стеганографічного аналізу тексту природньою мовою. Також, не слід забувати про наступний етап аналізу, а саме дискурс аналіз, ефективність роботи якого прямо пропорційно залежить від простоти та однозначності результатів морфологічно-синтаксичного дослідження тексту, а ефективне виявлення стеганографічних трансформацій, на етапі синтаксичного аналізу тексту залежить від гнучкості та універсальності обраного методу, який зможе підлаштуватися під різноманітні засоби стеганографії.

Хоча, як на етапі морфологічного, так і на етапі синтаксичного аналізу відбувається пошук статистичних закономірностей та зв'язків, визначаються невідповідності і робиться попередній висновок про можливу присутність контейнеру з прихованим повідомленням, проте, модуль дискурс аналізу виступає в головній ролі, оскільки саме дискурс аналіз може дати повну, цілісну картину для більш точних висновків про текст, а для цього треба отримати результати технічних перетворень, що проводяться під час морфологічного та синтаксичного аналізу повідомлення. Саме під час дискурс аналізу можна оцінити наскільки інформація осмислена, визначити чи був текст згенерований програмно, виділити теми та мікротеми тексту, що дозволяє значно скоротити початковий текст саме на смисловому рівні.

Аналітичний портал «Гуманітарні технології» в розділі «Дискурс-аналіз» [19] відзначає, що дискурс аналіз тексту являє собою єдність таких методів дослідження, як структурний, семіотичний, системний, символічний, риторичний, жанровий, аналіз наративу ключових слів, критичний та багатьох інших методів. Також виділяється 3 основні групи, основані на узагальненні цих різноманітних напрямів: текстуальний, інтертекстуальний та контекстуальний підхід [19].

Згадані підходи відносяться до процесу дискурс аналізу людиною, проте, якщо говорити за автоматизований комп'ютерний аналіз, особливо в області стегоаналізу, можна використовувати 2 підходи: текстуальний та інтертекстуальний. Так, Керолайн Спорледер в статті «Лексичні моделі для визначення немаркованих дискурсних відносин: чи допомагає WordNet?» [20] бере за основу інтертекстуальний підхід до автоматизованого дискурс аналізу, що полягає в дослідженні тексту в порівнянні зі схожими текстами шляхом гіперпосилань. Тому, такий підхід можливо використовувати в рамках задач стегоаналізу, таких, як виявлення наявності внесення змін до вже існуючого тексту шляхом пошуку початкового тексту та порівняння з ним. У випадку, коли йдеться про унікальне текстове повідомлення, написане людиною чи згенероване програмно слід використовувати текстуальний підхід, що дозволить на базі отриманих даних про нього на етапі морфологічно-лексичного дослідження дізнатися про наявність контейнеру з прихованим повідомленням.

Однією з важливих завдань дискурсного аналізатору є визначення основної теми тексту та його підтем, що досягається шляхом визначення ключових слів, побудови таблиць з частотою повторення синонімів. Такий підхід одночасно дасть змогу визначити чи були внесені зміни до тексту атакою заміни синонімів, а також дозволить зрозуміти чи є відхилення від основної тематики тексту та чи є текст осмисленим та логічним.

Крім того, як стверджує Бабіна О.І. в статті «Лінгвістична стеганографія: сучасні підходи. Частина 1» [21] дискурс аналіз може виявити незвичний розподіл синтаксичних типів та наводить приклад із текстом для аналізу, що являє собою інструкцію, де речення – це сукупність імперативів, а поява пасивної конструкцію буде виділятися зі структури дискурсу даного тексту [21 с.31]. Такий підхід зможе ефективно протидіяти багатьом методам автоматизованої стеганографії.

Як стверджує Єфременко Н.В. в статті «Лінгвістична стеганографія» [22], процес приховування повідомлення в текстових даних, як і у випадку з

аудіо чи візуальною інформацією зазвичай базується на надлишковості письмової мови [22, с.69]. Це та другорядна смислова інформація, що використовується для опису чи уточнення головної думки повідомлення. Тому її видалення забезпечить знищення і контейнеру з прихованими даними.

Усі згадані методи стеганографії мають один недолік, вони спотворюють основну думку тексту чи його логічну побудову. Якщо не брати до уваги механічні методи, що працюють з форматуванням тексту, які легко виявити, або навіть випадково знищити, то решта методів прямо впливають на смисл тексту.

Взяти, наприклад, підхід синонімічних підстановок, що описаний в статті «Лінгвістична стеганографія: сучасні підходи. Частина 2» Бабіної Ольги Іванівни [23]. Хоча автор стверджує, що при реалізації методу стоїть задача збереження характеристик статистичного розподілу синонімів замінників і доводить, що такий підхід забезпечує певну криптостійкість системи в результаті того, що при шифруванні однакової кількості біт повідомлення в тексті можуть бути використані різні слова субститути з одного класу еквівалентності [23, с.51]. Проте не зважаючи на криптографічну стійкість це відкриває можливість для атаки на стегосистему з допомогою дискурс аналізатору, який ефективно визначить, що використані синоніми спотворили зміст тексту.

Можна стверджувати, що автоматизований дискурс аналіз тексту охоплює усі відомі методи стеганографії та виводить стегоаналіз на більш високий та якісний рівень, що дозволить ефективно задовольнити основні потреби сучасної кібербезпеки в рамках протидії методам комп'ютерної лінгвістичної стеганографії.

В результаті оброблений текст буде максимально скороченим та позбавлений надлишковості без втрати основної думки, а стегоповідомлення спотворено чи втрачено.

**Висновок**

Для розв'язання задачі, направленої на виявлення шляхів задоволення потреб сучасної кібербезпеки в рамках протидії методам комп'ютерної лінгвістичної стеганографії було розглянуто найпопулярніші методи сучасної текстової стеганографії, доведено, що їх ефективність прямо залежить від наявності дієвих методів стегоаналізу, проаналізовано методи стегоаналізу текстової інформації. В результаті цього виявлено, що кібербезпека потребує проведення досліджень в напрямку розробки суто лінгвістичних методів стегоаналізу. Проаналізовано спроби сучасних дослідників удосконалити існуючі методи стегоаналізу тексту природньої мови та виявлена актуальність створення ефективних засобів автоматизованого комп'ютерного стегоаналізу. Виявлена потреба в аналізі саме осмисленості тексту, задовольнити яку можливо шляхом побудови текстового процесору. Проаналізовано класичну модель текстових процесорів, що складається з морфологічного, синтаксичного та лексичного блоків та виявлено необхідність її модифікації з метою використання блоку дискурсного аналізу.

Проведено аналіз поширених методів морфологічного аналізу та їх переваг, в результаті чого виявлено найбільш підходящий для потреб автоматизованого стегоаналізу текстової інформації метод логічного множення. Також було досліджено методи синтаксичного аналізу, серед яких імовірнісно-статистичний є тим, що задовольняє потреби стегоаналізу. Проаналізовано підходи до дискурсного аналізу, виявлено, що інтертекстуальний та текстуальний підхід можна використовувати саме для задоволення потреб комп'ютерного автоматизованого стегоаналізу, а їх використання в комплексі зумовить універсальність процесу дослідження тексту. Виявлено і інші переваги модуля дискурсного аналізатору, що дасть змогу перевіряти текстове повідомлення на осмисленість, виділяти його тему та підтеми і визначати ступінь їх логічної зв'язності.

Проведені дослідження підтверджують необхідність створення і розробку методу стиснення текстової інформації для лінгвістичної стеганографії, науково-практична реалізація якого дозволить задовольнити потреби сучасної кібербезпеки у виявлення та нейтралізації загроз, пов'язаних з дією засобів комп'ютерної лінгвістичної стеганографії та може бути використана, як підгрунтя для подальших наукових досліджень в області комп'ютерної лінгвістичної стеганографії в складі стегоаналітичних систем.

**Література**

[1] Bennett K., «Linguistic Steganography: Survey, Analysis, and Robustness Concerns for Hiding Information in Text», *CERIAS Tech. Report*, Purdue University, 30 p., May, 2004.

[2] Рябко Б., «Применение методов сжатия данных для непараметрического оценивания характеристик случайных процессов с дискретным временем», *Проблемы передачи информации*, № 4(43), с. 109–123., 2007.

[3] И. Нечта, «Эффективный метод стегоанализа базирующийся на сжатии данных», *Вестник СибГУТИ*, №1,с. 50-55, 2010.

[4] D. Roshidi; A. Faudziah; H. Hussain and others, «A Performance of Text Steganalytic System Using Genetic-Based Method», ARPN *Journal of Engineering and Applied Sciences*, №10(11), pp. 6216-6221, 2016.

[5] М. Мельник, «Повышение устойчивости стеганографической системы к атаке сжатием», дис. канд. техн. наук: 08.13.21, *ГВУЗ «Одесский национальный политехнический университет»*, 142 с., 2013.

[6] Н. Кошева, Н. Мазниченко, «Информационные технологии и защита информации в информационно-коммуникационных системах», *монография, под ред. В.С. Пономаренко, Харьковский нац. ун-т ім. Семена Кузнеца*, Вид. ТОВ «Щедра садиба плюс», 486 с., 2015.

[7] Н. Кухарська, О. Крижановська, «Аналіз методів текстової стеганографії», Електронний ресурс, Режим доступу: https://sci.ldubgd.edu.ua/handle/123456789/753

[8] В. Іванов, «Використання методів текстової стеганографії для захисту авторських прав в мережі Internet», *Проблеми інформатики та комп'ютерної тех-*

*ніки : пр. міжнар. наук.-практ. конф. (ПІКТ-2015)*, Чернівці, с. 171–173, 2015.

[9] Н. Ефременко, «Лингвистическая стеганография», *Вестник МГЛУ*, №619, с. 66-73, 2011.

[10] «Классификация методов стегоанализа», Електронний ресурс, Режим доступу: http://gr1g0ry.blogspot.com/2011/01/steganalysis-methods-classification.html.

[11] И. Нечта, «Применение статистического анализа для обнаружения скрытых сообщений в текстовых данных», *Вестник СибГУТИ*, №1, с. 23-29, 2012.

[12] И. Нечта, «Метод стегоанализа текстовых данных, основанный на использовании статистического анализа», *Вестник СибГУТИ*, №3, с. 27-34, 2011.

[13] А. Андреев, Д. Березкин, А. Брик, Ю. Кантонистов, «Вероятностный синтаксический анализатор для информационно-поисковой системы», Електронний ресурс, Режим доступу: http://www.inteltec.ru/publish/articles/textan/1kx5_9.shtml

[14] Л. Прокошенкова, «Дискурсивный анализ и его роль в современной лингвистике», *Вестник ЧГУ*, №4,с.451-456, 2006.

[15] О.Бабина «Корпусный метод автоматического морфологического анализа флективных языков», *Вестник ЮУрГУ*, №25, с.38-44, 2012.

[16] «Морфологический анализ, его виды», Електронний ресурс, Режим доступу: http://linguistics-konspect.org/?content=507 .

[17] П. Аношин, «Автоматический анализ текстов. Синтаксический и семантический анализ», Евразийский научный журнал, №6, 2017, Електронний ресурс, Режим доступу: http://journalpro.ru/articles/avtomaticheskiy-analiz-tekstov-intaksicheskiy-i-semanticheskiy-analiz/ .

[18] V. Kovar, «Automatic Syntactic Analysis for Real-World Applications», *PHD Thesis*, Masaryk University, Brno, 154 p., Spring 2014.

[19] А. Сарна, «Дискурс-анализ», Електронний ресурс, Режим доступу: http://gtmarket.ru/concepts/7232.

[20] C. Sporleder, «Lexical Models to Identify Unmarked Discourse Relations: Does WordNet help?», *Lexical-Semantic Resources in Automated Discourse Analysis,* №2(23), pp. 20-32.

[21] О. Бабина, «Лингвистическая стеганография: современные подходы. Часть 1», Вестник ЮУрГУ. Серия: Лингвистика, №3, с. 27-33, 2015.

[22] Н. Ефременко, «Лингвистическая стеганография», *Вестник МГЛУ*, №619, с. 66-73, 2011.

[23] О. Бабина, «Лингвистическая стеганография: современные подходы. Часть 2», *Вестник ЮУрГУ*. Серия: Лингвистика, №4, с.49-55, 2015.

**УДК 003.26 (045)**

***Федотова-Пивень И. Н., Тарасенко Я. В. Пути удовлетворения потребностей современной кибербезопасности в рамках противодействия методам компьютерной лингвистической стеганографии***

***Аннотация.*** *В статье проводится обзор распространенных методов текстовой стеганографии, а именно методов произвольного интервала, синтаксических и семантических методов, исследуются существующие пути противодействия им, а также средства автоматизированного лингвистического анализа текста (морфологического, синтаксического, дискурсного) для автоматизации текстового стегоанализа. Проводится обобщение существующих методов и отмечаются такие, что могут получить дальнейшее развитие и усовершенствование при решении задач противодействия средствам компьютерной линевистической стеганографии и могут быть реализованы посредством лингвистического процессора (метод логического умножения для морфологического, вероятностно-статистический метод для синтаксического и текстуально-интертекстуальный подход для дискурсного анализа). Таким образом, доказывается необходимость проведения научного исследования и создания метода сжатия текстовой информации для лингвистической стеганографии.*

***Ключевые слова:*** *кибербезопасность, методы произвольного интервала, семантические методы, морфологический анализ, синтаксический анализ, дискурсивный анализ, лингвистический процессор, автоматизированный стегоанализ.*

***Fedotova-Piven I., Tarasenko Ya. The ways of relevant of modern cybersecurity in context of countering the methods of computer linguistic steganography***

***Abstract.*** *The article reviews the common methods of textual steganography, such as methods of random intervals, syntactic and semantic methods, detects the existing ways of confronting them and also reviews the methods of automated linguistic analysis of the text (morphological, syntactic, discourse) for the purpose of the textual steganalysis automation. A generalization of existing methods is carried out and those that can acquire further development and improvement in solving the problem of counteracting the means of computer linguistic steganography, and can be implemented using a linguistic processor (the logical multiplication method for morphological, probabilistic-statistical method for syntactic and textual-intertextual approach for discursive analysis) are noted. Thus, the necessity of carrying out the scientific research and creating a method of the textual information compression for linguistic steganography is proved.*

***Key words****: cyber security, linguistic steganography, random interval methods, syntactic methods, semantic methods, morphological analysis, parsing, discursive analysis, linguistic processor, automated steganalysis.*

Отримано *10 жовтня 2017 року*, затверджено редколегією *30 жовтня 2017 року*