УДК 004.056.5:004.7

П.В. Сироватка, М.С. Літош, Н.І. Довгич,
К.П. Ануфрієнко, кандидат технічних наук НАУ, асистент
(Національний авіаційний університет, Україна)

ПІДВИЩЕННЯ РІВНЯ СТЕГАНОГРАФІЧНОЇ СТІЙКОСТІ ДО СУБ'ЄКТИВНИХ АТАК

У роботі розглядаються сучасні методи лінгвістичної стеганографії, аналізуються їх переваги та недоліки. Встановлюється, що суб'єктивні атаки суттєво впливають на рівень стеганографічної стійкості системи. Пропонується використання алгоритму синонімічних замін поряд з методом прихованих або замаскованих шифрів.

Серед усього спектру систем захисту інформації від несанкціонованого доступу особливе місце займають стеганографічні системи. На відміну від інших систем, вони опираються не лише на властивості самої інформації, а й на властивості її матеріальних носіїв, особливості вузлів її обробки, передачі й зберігання.

У даний час важко переоцінити необхідність приховування наявності самого факту передачі інформації, що захищається у діяльності більшості як державних, так і комерційних установ. Адже, криптографічні засоби на даний час можуть лише затягнути процес дешифрування даних зловмисником, але навряд чи зупинять його. Операції з важливими даними завжди пов'язані з підвищеним ризиком, особливо якщо інформація надзвичайно важлива, наприклад, ведення секретних переговорів, передача номерів рахунків, кодів доступу і т.п.

Питаннями захисту конфіденційної інформації займається стеганографія – наука про приховування інформації шляхом збереження в таємниці самого факту передачі [1]. Тобто приховування одних повідомлень в інших повідомленнях. У якості контейнера (звичайного повідомлення, придатного для вбудовуваної інформації, що захищається) використовуються типові документи, мультимедійні дані та звичайний текст.

Стеганографічні методи захисту інформації в останні роки розвиваються досить активно, постійно з'являються нові методи приховування інформації і нові методи стеганоаналізу. Зростає і загальне число як зарубіжних, так і вітчизняних публікацій. Разом з тим, аналіз останніх дозволяє говорити про те, що переважна більшість досліджень [2, 3] спрямована на мультимедійний контент. Хоча на даний момент величезна кількість інформації представлена в текстовому вигляді: книги, статті, електронне листування, документи, звіти і багато іншого, і всі ці матеріали можуть бути ефективно використані в якості контейнерів для прихованої передачі інформації [3]. Не дарма величезною популярністю в процесі приховування інформації зараз користується лінгвістична стеганографія, яка дозволяє приховати кодовану довільну інформацію у тексти, спираючись на особливості мови та лінгвістичні ресурси.

Розглядаючи роботи зарубіжних фахівців [1-3], присвячені текстовій та лінгвістичній стеганографії, можна помітити, що автори цих робіт чітко розмежовують методи та алгоритми стеганографії із захисту інформації, що приховується від "роботів" і від людей. Перші спрямовані на захист інформації при тотальному скануванні всієї кореспонденції програмним пошуковими роботами і аналізаторами, другі спрямовані на захист інформації при уважному перегляді тексту людиною.

Аналіз найбільш популярних стеганографічних програм, які дозволяють вбудовувати інформацію в текстові файли (FFENCODE, SecureEngine, Центуріон), а також аналіз статистичних даних [4] вказують на нестійкість таких стегосистем (таємність забезпечується за рахунок збереження в таємниці алгоритму вбудовування інформації, що суперечить принципу Кергофа). Звідси виникає необхідність створення лінгвістичної стегосистеми, яка б була стійкою до атак стегоаналітиків (противників).

На сьогодні відомо декілька методів лінгвістичної стеганографії [3] (див. рис. 1).

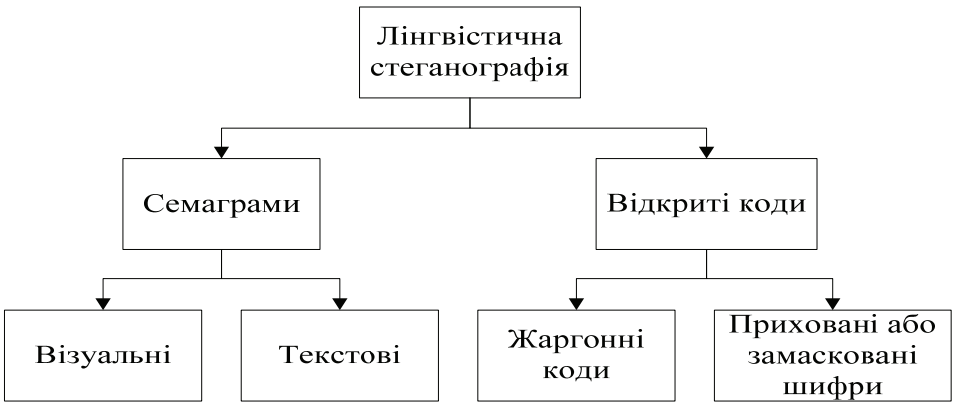

Рис.1. Класифікація методів лінгвістичної стеганографії

Семаграми приховують інформацію за допомогою символів або знаків. *Візуальна семаграма* використовує нешкідливі на перший погляд чи звичайні фізичні об'єкти для передачі повідомлення, наприклад, “кривульки” або розташування елементів на робочому столі або веб-сайті. *Текстова семаграма* приховує повідомлення, змінюючи зовнішній вигляд тексту-контейнера, наприклад, ледь помітні зміни в розмірі або типі шрифту, додаючи додаткові пробіли або різні завитки в буквах або рукописному тексті.

Відкриті коди приховують повідомлення в “законному” повідомленні-контейнері такими способами, які не видимі для непідозрюваного спостерігача. Таке повідомлення-контейнер іноді називають відкритою комунікацією, тоді як приховане повідомлення – таємною комунікацією. Ця категорія підрозділена на жаргонні коди і приховані або замасковані шифри. Жаргонний код, як і передбачає назва, використовує мову, яка зрозуміла одній лише групі людей, але не має сенсу для інших. *Жаргонні коди* включають в себе нанесення піктограм (символи, використовувані для вказівки присутності та типу сигналу бездротової мережі), таємну термінологію, або невинну розмову, яка передає особливий сенс внаслідок того, що факти відомі тільки співрозмовникам. *Приховані чи замасковані шифри* приховують повідомлення в носії-контейнері так, щоб його міг відновити будь-який, хто знає секрет того, як воно було приховано.

Існуюча на сьогоднішній день велика різноманітність різних форматів зберігання та подання електронних документів надає широкі можливості для побудови на їх основі систем прихованої передачі інформації. Разом з тим більшість стеганографічних методів виявляються нестійкими у разі збереження електронного документа-контейнера в іншому форматі. І також більшість з цих методів легко піддаються суб’єктивним атакам зловмисників. При цих атаках [4] зловмисник уважно розглядає стегосистему, намагаючись визначити “на око”, чи є в ній приховане повідомлення. Ясно, що подібна атака може бути проведена лише проти абсолютно незахищених стегосистем. Тим не менш, вона, напевно, найбільш поширена на практиці, принаймні, на початковому етапі розкриття стегосистеми. Первинний аналіз також може включати в себе наступні заходи:

1. Первинне сортування стего за зовнішніми ознаками.
2. Виділення стего з відомим алгоритмом вбудовування.
3. Визначення використаних стегоалгоритмів.
4. Перевірка достатності обсягу матеріалу для стегоаналіза.
5. Перевірка можливості проведення аналізу за окремими випадками.
6. Аналітична розробка стегоматеріалів. Розробка методів розкриття стегосистеми.
7. Виділення стего з відомими алгоритмами вбудовування, але невідомими ключами.

Також, варто зазначити, що існують ще й синтаксичні та семантичні методи лінгвістичної стеганографії, які ґрунтуються на зміні пунктуації, стилю і структури тексту, що є неприйнятним в багатьох сферах суспільного життя. Для людини такі тексти виглядають повною нісенітницею з граматичними помилками, що є неприйнятним. Також і розробка

програмних реалізацій даних методів для української та російської мови викликає ряд труднощів, оскільки повинна зберігатися осмисленість і однозначність тексту документа-контейнера.

На фоні цих методів найбільш стійкими відносно суб'єктивних атак виступають приховані або замасковані шифри. Але й вони мають свої певні недоліки – проблеми, які виникають при формуванні стеганограм за певними правилами, в яких є важливим порядок розміщення відповідних букв, символів, цифр на відповідних місцях в контейнері.

Тому, *метою* даної роботи є підвищення рівня стеганографічної стійкості до суб'єктивних атак, шляхом використання особливостей лінгвістичної стеганографії.

Ситуацію можна було б покращити використовуючи паралельно з класичними методами прихованих або замаскованих шифрів лінгвістичної стеганографії – алгоритму синонімічних замін. Розглянемо його детальніше [2].

Вхідними даними в пропонований алгоритм служить двійкова інформація, призначена для шифрування, і що має текст, за обсягом приблизно в 200 разів більший, ніж у шифруємої інформації. Формат тексту довільний, але він орфографічно і синтаксично правильний, щоб не спровокувати виправлень при передачі. Послідовності цифр або особистих імен допускаються, але вони збільшують необхідну довжину тексту. Алгоритм включає наступні кроки.

Пошук синонімічних слів. У тексті відшукуються слова і багатослівні вираження, що мають синоніми. Якщо одночасно знайдена послідовність слів і її підпослідовність, перевага віддається більшій.

Формування об'єднаних синонімічних груп. Послідовно розглядаються синонімічні слова тексту. Якщо у відповідній синонімічної групі тільки абсолютні синоніми, вона приймається беззастережно. Якщо в групі є хоч один відносний синонім, всі такі синоніми піддаються операції транзитивного замикання. При замиканні для кожного синоніма перевіряється, чи не є він членом будь-якої іншої синонімічної групи. Якщо це так, додаткова група приєднується до початкової без повторів. Далі приєднані синоніми проглядаються на приналежність до інших, ще не розглянутих синонімічних груп, і так до вичерпання поповнень. Замикання здійснюється також через омоніми. Аналізується, чи не є омонімічне вихідне текстове слово або який-небудь член його синонімічної групи. Якщо це так, і якщо ще не розглянутий омонім має синоніми, залучається група синонімів цього омоніму. Кожна знову залучена група використовується для розширення і т.д. Процес кінцевий, але іноді дає велику об'єднану групу. Транзитивне замикання необхідно, оскільки робить складу об'єднаної групи не залежать від того, з якого члена замикання починається.

Перевірка словосполучень. Якщо група містить лише абсолютні синоніми, вона не перевіряється на контекст, а для перевірки на сполучуваність з нею інших слів може братися будь-який її член. Якщо ж група має відносні синоніми s_i, вони підлягають перевірці на сумісність із зовнішніми повнозначними словами w_j ліворуч і праворуч від об'єктів аудиту групи. Якщо зовнішнє слово w_i не синонімічне або має лише абсолютні синоніми, то перевіряється, з якими з s_i воно утворює однотипні словосполучення. Ті s_i, які не формують словосполучень з w_j, відкидаються. Якщо зовнішнє слово w_j саме належить об'єднаній групі з елементами w_{jk}, то перевіряються всі однотипні словосполучення пар $\{w_{jk}, s_i\}$ при всіх i і k. Відсутність словосполучення хоча б з одним зовнішнім елементом веде до відкидання перевірки елемента.

Кодування. Послідовність профільтрованих груп сканується. Якщо їх розміри кратні степеню двійки або скорочені до найближчого степеня двійки, для чергової групи довжиною 2^n з закодованого повідомлення виділяється склад довжини n і його двійкове значення береться як внутрішньогруповий номер синоніма і підставляється в текст замість вихідного. Та ж операція повторюється для всіх груп уздовж тексту. Якщо є групи, по довжині не рівні ступеню двійки, всі довжини груп перемножуються і береться ступінь 2^N, найближчі до отриманого добутку вниз. Потім від кодованої інформації відсікається склад довжини N і з нього шляхом послідовних поділів та знаходжень залишків знаходяться номери омонімів для заміни синонімічних слів у тексті. Якщо текстовий синонім при кодуванні виявився

заміненим, то в загальному випадку перевизначаться морфосинтаксичні характеристики замінника і контексту.

При цьому така система формування стеганограми дозволяє зберегти зовнішню “безневинність” і осмисленість тексту-контейнеру [5].

Також є доцільним і шифрування даних перед відправкою певним криптографічним алгоритмом, що безумовно підвищить рівень стійкості. Створену систему можна буде використовувати в багатьох сферах суспільного життя, таких як банківська, фінансова, оборонна галузь та інші галузі науки та техніки. Вона зможе забезпечити гарантовану передачу цифрових даних будь-якими каналами, з мінімальною можливістю злому, адже на відміну від багатьох існуючих стеганографічних систем, які не можуть протистояти атаці стегоаналітика, ця система буде проходити відповідну обробку створеної стеганограми засобами синонімічних словників.

Висновки

Використання алгоритму синонімічних замін поряд зі звичайними методами лінгвістичної стеганографії значно підвищить стеганографічну стійкість стеганограм перед суб’єктивними атаками зловмисників, завдяки тому що така система формування стеганограми дозволяє зберегти зовнішню “безневинність” і осмисленість тексту-контейнеру. Створена система зможе протистояти тотальному скануванню програмними пошуковими роботами і аналізаторами, а також уважному перегляду тексту людиною. Вона зможе забезпечити гарантовану передачу цифрових даних будь-якими каналами, з мінімальною можливістю злому.

Список літератури

1. Большаков И. А. Два метода синонимического перефразирования в лингвистической стеганографии. / И.А. Большаков // Центр Компьютерных Исследований. – 2007. – № 1. – 143 с.

2. Мельников Ю. Банковские технологии / Мельников Ю., Колошеин Ю. – М. : Наука, 2003. – С. 35-37.

3. Кесслер Г. Стеганография для судебного исследователя. Краткий Обзор. (Перевод: Капинус О.В., Михайлов И.Ю., Бочков Д.С.) / Гари Кесслер // Компьютерно-техническая экспертиза. – 2008. – №2. – С. 13-25.

4. *Мельчук И.А*. Опыт теории лингвистических моделей “Смысл=Текст”: Семантика, синтаксис / И.А. Мельчук. – М. : Наука, 1974. – 314 с.

5. Большаков И.А. Тезаурус в системах подготовки текстов: каким ему быть? / И.А. Большаков // Междунар. форум по информ. и джум. – 1991. – Т №2. – С. 31.