УДК 004.415.24:004.056.5

АНАЛІЗ МЕТОДІВ ТЕКСТОВОЇ СТЕГАНОГРАФІЇ

Крижановська О.Л.

Кухарська Н.П., канд. фіз.-мат. наук, доцент

Львівський державний університет безпеки життєдіяльності

Текстові документи, незважаючи на строгість їх оформлення і неможливість модифікації окремих бітів, є одним із застосовуваних у стеганографії видів контейнерів. Сучасні спеціалізовані текстові процесори надають користувачам доволі широкі можливості переддрукової підготовки текстових документів. Це і форматування окремих символів, абзаців, сторінок, розділів документа, використання різних шрифтів, всеможливих списків, вставлень, а також відображення безпосередньо в самому документі, під час попереднього перегляду, таблиць, рисунків, різних діаграм і графіків. Звісно, таке різноманіття можливостей спричиняє складну структуру файлів текстових документів. Розбиття документа на блоки, нетривіальні зв'язки між різними його фрагментами, різноманітність у межах одного файлу типів даних (текст, таблиці, графіки, рисунки, анімації та ін.), безліч різних полів і спеціальних заголовків, наявність великих обсягів невідображуваної службової інформації сприяє використанню текстових документів у цілях стеганографії.

Виділяють три групи методів приховування інформації з обмеженим доступом у текстових документах [1]:

- методи довільного інтервалу;
- синтаксичні методи;
- семантичні методи.

Методи довільного інтервалу в деяких випадках показують досить непогані результати. По-перше, зміна кількості пропусків, наприклад, у кінці текстового рядка не викликає жодних змін у значенні фрази або речення. По-друге, непосвячений читач сумнівно чи помітить незначні модифікації вільних, наявних і до застосовування процедури вбудовування, місць. До методів цієї групи належать, зокрема:

1. *Метод зміни кількості пропусків між реченнями.* Цей метод дозволяє вбудовувати в текстовий файл секретне повідомлення, проставляючи один (що відповідає приховуваному біту "1") або два (їм ставиться у відповідність приховуваний біт "0") пропуски після кожного символу кінця речення. Описаний метод має ряд недоліків: для вбудовування навіть незначної кількості бітів потрібен текст значного розміру; можливість приховування залежить від структури текстового контейнера (в окремих випадках у текстових документах можуть бути відсутні знаки кінця рядка; деякі текстові редактори можуть автоматично додавати після крапки пропуски).

2. *Метод зміни кількості пропусків у кінці текстових рядків.* Кількість біт, що можна приховати у тому чи іншому рядку, визначається різницею між кількістю символів у найдовшому

рядку тексту і кількістю символів у поточному рядку. При цьому, якщо приховується біт “0”, до стеганограми дописується звичайний пропуск (значення ASCII-коду 32), а якщо біт “1” – нерозривний пропуск (значення ASCII-коду 160). Такий підхід дає можливість дещо збільшити, порівняно з попереднім методом, кількість інформації, яку можна приховати у документі. Цей метод може бути застосований до будь-якого тексту, при чому зміни у форматі останнього будуть візуально непомітними. Недоліком методу є те, що деякі програми обробки тексту можуть ненавмисно вилучати зайві пропуски.

3. *Метод зміни кількості пропусків між словами вирівняного за шириною тексту.* Відповідно до алгоритму цього методу біти конфіденційних даних вбудовуються у текстовий документ на основі керованого вибору позицій, куди розміщуватимуться додаткові пропуски. Один біт секретних даних вбудовується в контейнер шляхом модифікації пари пропусків, що охоплюють одне слово речення. У залежності від значення приховуваного біту, один пропуск додається на початок або в кінець слова. Зазвичай цей метод дає можливість вбудовувати по кілька біт в один рядок. Недоліком є те, що через обмеження, які накладаються вирівнюванням тексту за шириною, не кожен пропуск між словами може бути використаний для вбудовування даних.

Такі операції як форматування, заміна символів табуляції пропусками, вилучення зайвих пропусків в кінці рядків і т.д., можуть призвести до псування або до повного знищення конфіденційного повідомлення, прихованого методами довільного інтервалу. Значно більшу стійкість до подібних спотворень мають інші методи, що оперують безпосередньо самим текстом, його реченнями і словами.

Синтаксичні методи полягають в зміні пунктуації, абревіатури і скорочень тексту. Незважаючи на те, що розстановка знаків пунктуації є достатньо строго обумовленою правилами використовуваної мови, існують випадки, коли ці правила виявляються неоднозначними або ж відхилення від них не веде до суттєвого спотворення смислу тексту. Наведемо приклад:

“Використовують червоний, синій, зелений кольори”.

“Використовують червоний, синій і зелений кольори”.

Одна із ком першого речення була замінена у другому сполучником “і”. У результаті ми отримали два варіанти одного речення. Першому варіанту можна поставити у відповідність біт “0” приховуваного повідомлення, а другому – “1”. Абсолютно аналогічно можуть бути використанні скорочення і абревіатури. До синтаксичних методів відносять також методи, що ґрунтуються на заміні стилю і структури речення без помітного спотворення вихідного смислового навантаження. Синтаксичні методи слід використовувати з ретельною обачністю, бо внесені ними зміни можуть призвести до зниження ефекту сприйняття тексту, надати йому протилежного змісту чи привернути увагу цензора.

Семантичні методи є найбільш привабливим напрямком в текстовій стеганографії. Вони відзначаються високою ефективністю, так як застосовують різні підходи маніпулювання безпосередньо самими реченнями і словами, а не другорядними елементами чи незначними особливостями текстів. Методи, що відносяться до даного напрямку, ґрунтуються на використанні синонімів. Майже у будь-якому достатньо довгому реченні зустрічаються слова, котрі без втрати змісту можуть бути замінені синонімами. Якщо для якогось слова існує не один, а декілька синонімів, то формуються спеціальні таблиці замін. У таких таблицях кожному синоніму ставиться у відповідність код, що містить декілька приховуваних двійкових символів. Наприклад, слову “проте” може бути поставлено у відповідність біт “0”, а слову “однак” – “1”. Водночас, слід зауважити, у ряді випадків використання семантичних методів ускладнюється деякими нюансами вживання ключових слів у реченнях, їх смисловими відтінками.

ЛІТЕРАТУРА

1. Конахович Г.Ф. Компьютерная стеганография. Теория и практика / Г.Ф. Конахович, А.Ю. Пузыренко. – К. : Изд-во "МК-Пресс", 2006. – 288 с.