Annual of Konstantin Preslavsky University of Shumen
vol. XX C, 2019, pp. 67 - 92

TEXT STEGANOGRAPHY METHODS*

TEODORA T. STOYANOVA, STANIMIR K. ZHELEZOV

ABSTRACT: *Nowadays, the information protection is a highly topical issue in a number of areas. Steganography is a scientific field of application, a set of technical skills and the art of the ways to hide the fact of transmitting (availability) of information. The most common steganographic methods are reviewed and classified. The principles of the textual steganography are reviewed. A classification of the methods of textual steganography is made.*

KEYWORDS: *Steganography; Text steganography; Stego methods; Information hiding; Information security.*

2010 Math. Subject Classification: *94A99, 68P20, 68P2530*

МЕТОДИ НА ТЕКСТОВАТА СТЕГАНОГРАФИЯ†

ТЕОДОРА Т. СТОЯНОВА, СТАНИМИР К. ЖЕЛЕЗОВ

1 Компютърна стеганография

В наше време защитата на информацията е много актуална в редица области. Защитата в Интернет е от голямо значение и за бизнеса, и за държавата. Банкови карти, разплащателни сметки, интернет банкиране, пазаруване от всяка точка на света от различни мобилни устройства – всичко това създава опасност от

* This paper is (partially) supported by Scientific Research Grant № RD-08-96/01.02.2019 of Konstantin Preslavsky University of Shumen

† Статията е частично финансирана по проект № РД- 08-96/01.02.2019 "Защита и надеждност на данни във виртуални и web среди, графични файлове, 3D моделиране на терени" на ШУ

злоупотреби. Широко разпространение напоследък получи използването на стеганографски методи за скриване и предаване на конфиденциална информация. Скриването на факта за предаване на информация е добър начин за предотвратяване на атаки [1].

Стеганография (steganography) е научно - приложна област, съвкупност от технически умения и изкуство за начините за скриване на факта на предаване (наличие) на информация [2]. От няколко години се използва и терминът стеганология (steganology), обхващащ два смислово противоположни компонента - стеганография и стеганализ. Стеганализът (steganalysis) представлява съвкупност от методи и технологии за откриване на секретни комуникации, които използват стеганографски методи [2,3].

Най-общо в състава на една стеганографска система се включват секретно съобщение, контейнер, стегоключ, стегометод и канал за предаване на данни.

Реализацията на стеганографските методи е представена на Фиг. 1. Тя се използва при всяка стеганографска комуникация, независимо от конкретно използвания метод. В началото на схемата винаги стои изпращач, който иска да скрие съобщение, така че то да остане неразбрано от всички останали с изключение на човека получател, за когото е предназначено то [2].

Фиг. 1. Реализация на стеганографските методи

2 Класификация на най-разпространените стеганографски методи

Главната цел в стеганографията е да се създаде надежден начин за вграждане и извличане на данни без това да предизвика подозрение. Под данни може да се разбира всякаква информация: текст, съобщение, изображение и др., като различните видове стегосистеми скриват тези данни в различни видове носещи файлови формати. Основна цел на стеганографията е да осигури конфиденциалността на вградената информация чрез скриване на нейното съществуване.

Според съвременните схващания за стеганография тя се дели на високотехнологична и класическа.

Терминът класическа стеганография се използва само за да се формулира съвкупността от огромния брой исторически развили се методи, системи, техники, приложения и др. за скриване на факта на съществуване на съобщения и комуникации, без използване на съвременни високотехнологични способи [1,4].

Могат да бъдат изброени редица нискотехнологични стеганографски техники:

- запис на съобщение върху страничните страни на колоди от карти, подредени в условен предварително уговорен ред (след това картите се предават разбъркани);
- съставяне на съобщения чрез пробиване на дупчици с игли на букви от печатен текст в определено издание (думите се отделят с дупки между буквите);
- писмо чрез възли на конци, където всяка буква се кодира чрез различна дължина в сантиметри (напр. А - 1 см, Б - 2 см и т.н.);
- надписи на обратната страна на етикети на бутилки, флакони, буркани и др.;
- текст под залепена пощенска марка;
- акростихове и други езикови игри;
- използване на "развалена пишеща" машина, в която някои букви се печатат по-високо или по-ниско от реда (вземат се

предвид реда и броя на тези букви, а също така и междините при техните появявания);

- ръчен запис на ноти в нотна тетрадка (нотите имат значения според Морзовата азбука или друг код);
- запис във вид на кардиограма или график на технологичен процес (пак Морзова азбука - върховете на графика са точки, а тези по-ниско - тирета, и др.);
- използване на симпатични мастила;
- микроточка и др.

Класическата стеганография не е подходяща във всички ситуации или по-скоро има някои недостатъци и това е причината високотехнологичните методи да са за предпочитане:

- Бавно изпълними процедури по кодиране, декодиране и транспорт на съобщението;
- Информационният носител обикновено е лесно разрушим, лесно може да бъде компрометиран и дори загубен;
- Ограничени количества материал, това включва информационни носители и материали необходими за създаването им и това на материалите необходими за скриване и разкриване на съобщения;
- Информационните носители често са обемни и не са удобни за съхраняване за дълъг период от време;

Високотехнологичната стеганография е термин, използван от някои автори за обобщаване на направленията за скриване на съобщения с използване на комуникационните и компютърни технологии, нанотехнологиите и съвременните постижения на биологията [2].

Стегометодите позволяват скриване на данни в различни контейнери: текстови документи (електронни статии, книги, писма) в графични файлове (рисунки, банери, фотографии), видеофайлове (клипове, филми, анимация), в звукови файлове (музикални произведения, реч, природни звуци), в кода на HTML-страници, в субтитрите на филми, в съобщения, предавани с помощта на SMS, MMS, чат, блогове, и др. Текстови съобщения

могат да бъдат скрити в неизползваните области на Flash-паметите, твърдите и оптическите дискове. Като се има в предвид, че всеки вид контейнер има различни формати, а за скриване на информацията могат да се използват разнообразни методи, то се вижда колко многомерни са стеганографските задачи [2,5,6].

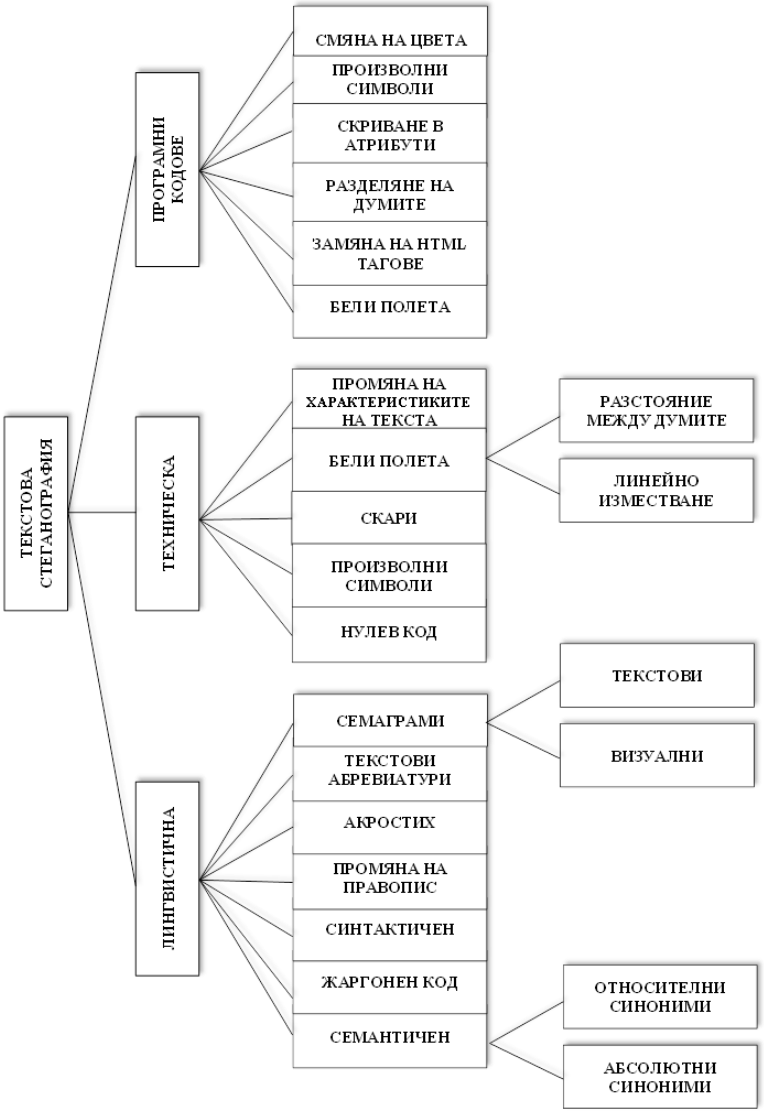
3 Текстова стеганография

Текстовата стеганография може да включва всичко - от промяна на форматирането на съществуващия текст, до промяна на думите в текста, да генерира произволни поредици от символи и т. н. [7]. Текстовата стеганография се смята за най-трудната поради дефицита на излишна информация, която е налична в останалите мултимедийни файлове. Структурата на текстовите документи е идентична с това, което наблюдаваме, докато в други видове документи, като изображения, структурата на документа е различна от това, което наблюдаваме. Следователно, в тези документи, може да се скрие информация чрез въвеждане на промени в структурата на документа, без да се прави осезаема промяна в крайната визуализация. Текстовия файл изисква по-малко памет при съхраняване и по-бързо, както и по-лесно предаване. Това го прави предпочитан пред другите видове стеганографски методи начин за предаване на скрита информация [8].

Текстовата стеганография използва за контейнер текстови файлове. Тя има предистория от времето на пишещите машини. През осемдесетте години на XX век, за да проследи изтичането на информация от „Даунингстрийт 10” към пресата, британския министър-председател Маргарет Тачър въвежда специални текстообработващи техники, кодиращи името на оторизирания получател на документа в интервалите между думите, така че лицата, отговорни за изтичане на информация да могат да бъдат идентифицирани [1,9].

В днешно време интензивността на потоците на изцяло текстова информация в комуникационните канали за връзки в Интернет непрекъснато се увеличава. Високата интензивност на текстовия трафик дава възможност за предаване на секретни съобщения чрез директното им поставяне в съществуващите текстове, макар и с неголяма скорост, но оставяйки ги незабелязани.

На Фиг. 2 е показана класификация на видовете текстовата стеганография и нейните методи. Тя се разделя на три главни поднива – лингвистична, техническа и в програмен код.

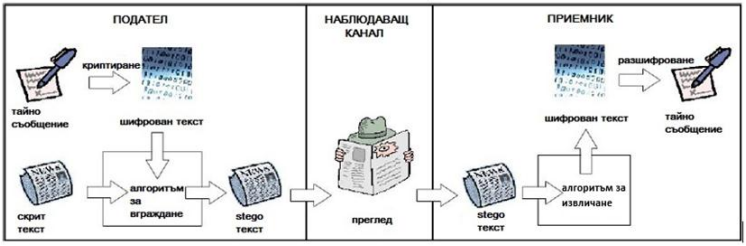


Фиг. 2. Класификация на видовете текстова стеганография

3.1. Лингвистична стеганография

Лингвистичната стеганография е също дял от класическата стеганография, но при развитието на съвременните технологии, нейните стеганографски методи се реализират основно с компютърни системи и софтуерни продукти. Поради тази причина подвидовете на класическата лингвистична стеганография се запазват, като към средствата за тяхната реализация се добавят софтуерни продукти и компютърни системи.

При лингвистичната стеганография е важно да бъде съхранен съществуващия текст по смисъл и съдържание. По този начин всеки „безобиден“ текст, който не привлича внимание с външния си вид - формат, шрифт, правопис, морфология, синтаксис и лексика, може да бъде носител на скрита информация. Всички тези черти трябва да се отнасят единствено и само към темата на текста (независимо от това как е представен и на какъв носител се пренася).



Фиг. 3. Структура на лингвистичната стеганография

- **Семантичен метод**

Този метод използва синоним на думата за скриване на данни, но понякога това може да промени действителното значение на текстовия файл [10]. Определят се два синонима, които отговарят за значението на скрития бит. Примерно, съюзът „но” може да бъде използван за носител на 0 бита, а словосъчетанието „би могло”, като носител на 1 бит.

За използване на семантичния метод за скриване на информация е необходима таблица на синонимите, показана в Таблица 1 [8]. При това трябва да се отчита възможността за повече от един синоним на една дума.

Таблица 1

Big	Large
Small	Litlle
Chilly	Cool
Smart	Clever
Spaced	Stretched

Таблица на синонимите

Създадените синонимни речници за целите на семантичния метод, трябва да са пълни и адекватни за дадения език. Цялата лексика на даден език се разделя на множество групи в различен обем. В рамките на групите думите са сходни, както граматически, така и семантично. В тези речници често се включват по широк обзор на синонимите, отколкото традиционно значимите за даден език от гледна точка на лексикологията. WordNet може да се използва за автоматично генериране на синонимни таблици. Проблем възниква, когато използването на синоним променя значението на кодираните данни. Например, възниква проблем с избора на двойката синоними „cool“ и „chilly“. Да наречеш някой „cool“ има много по-различно значение от „chilly“ [11].

Синонимите се разглеждат като абсолютни и относителни, еднословни и многословни.

Абсолютни синоними са тези, които са определени от лексикологията, например любов и обич, красив и хубав, различен и нееднакъв и т.н. Относителните синоними се заменят само в контекста на даден текст, примерно САЩ с Америка, ОНД с Русия и т.н. За разлика от относителните синоними,

абсолютните могат да се прилагат към синонимни заместванията, независимо от контекста.

Многословните синоними са например електро ток (от електрически ток), като в традиционните синонимни речници обикновено не са включени, но за този аспект на стеганографската наука са от значение. Друг пример е съединяването на две думи, като в речниците те са малко, но в говоримия език техният брой се увеличава и те често се използват в едни и същи текстове. Тук се използват и синоними, които имат едно и също значение с повече думи, примерно: "секретен“, “таен“, “скрит“, “конфиденциален“, “негласен“, “неизвестен“, “засекретен“, „закрит“ и дават възможност да се скрият повече битове.

Тук създадените методи се базират на синонимно перифразиране и запазване на текстовия смисъл, като по този начин се подсигурява надеждност и безопасност на вложеното съобщение.

- **Метод на жаргонния код**

В метода на жаргонния код се използва език, който се разбира от една група хора, но е безсмислен за други. Жаргонните кодове включват различни условности (терминология или невинен разговор), който предава специално послание. Тези условности предварително са известни само на хората, които трябва да получат посланието.

Този метод вероятно е най-очевидната форма на лингвистична стеганография. Съобщение, което е кодирано в много отношения прилича на заместващ шифър, но вместо да се заменят отделни букви, се променят самите думи.

- **Синтактичен метод**

Чрез поставянето на някои препинателни знаци, като например точка (.) или запетая (,) на подходящи места, може да се скрие информация в текстов файл. Този метод изисква да се

идентифицират подходящите места за поставяне на препинателни знаци [12].

Писменият език предоставя достатъчно възможности за синтактическо скриване на данни, тези възможности не се наблюдават в класическите произведения. Това е така, защото правилата за пунктуация се считат за неоднозначни, и противоречивото им използване, може да стане обект на внимание за редактора. Такива случаи са възможни, когато изменението на пунктуацията води до снижаване на възприемчивостта на текста или до предаване на текста, на съвсем различен смисъл. Затова синтактичния метод трябва да се използва много внимателно.

Към синтактическият метод се отнася и метода за изменение на стила и структурата на текста, без значително изменение на заложения смисъл. Например, изречението „съществуват не малко случаи, когато правилата на пунктуацията се явяват неоднозначни“ може да се формулира и като „правилата на пунктуацията се явяват неоднозначни в много случаи“. Такива методи се явяват още по незабележими за нарушители, в сравнение с методите за изменение на пунктуацията. Възможно е тяхното използване да не се подава на анализ на компютърна автоматизирана стеганографска система.

Ортографическият метод за скриване на информация в текста се осъществява с помощта на влагане на грешки в текста. Грешките се разпределят в текста в съответствие с ключа, който определя думите, за които е нужно да се проверяват ортографически. При наличие на грешка се счита, че кода се равнява на 1 бит, а при отсъствие 0 бита.

- **Метод на промяната в правописа**

Този метод използва едни и същи думи, които са написани по един начин на британски и американски английски [8,11]. На английски някои думи имат различен правопис, така че можем да скрием данни в текста, като заместваме тези думи показани на

Таблица 2. Този метод е съставен от две части, едната е скриваща програма, която отговаря за скриването на данни в текст. Друга е програма за извличане, която извлича данни от текста, съдържащ скрити данни. Отначало се подготвя списък, съдържащ думите, които имат различен правопис във Великобритания и САЩ. Методът на скриване търси съществуващи думи от списъка в текста, като при поставяне на дума от първата колона в изречението се скрива 0 бита, а от втората – 1 бит. По този начин данните ще бъдат скрити в съответния текст. Този метод има малък капацитет за скриване на данни в текста. Това е свързано с основния текст и неговия размер, но като цяло капацитетът му е много малък.

Таблица 2

American Spelling	**British Spelling**
Favorite	Favourite
Criticize	Criticise
Fulfill	Fulfil
Center	Centre
Dialog	Dialogue
Medieval	Mediaeval
Check	Cheque
Defense	Defence
Tire	Tyre

Промяна в правописа

- **Метод на акростиха**

Една разновидност на лингвистичната стеганография, наричана акростих, е била една от най популярните древни стеганографски техники. Този метод е използван и през XX век в Първата световна война и от германците, и от съюзниците. Един от известните примери е в следващите стихове на руският поет Николай Гумильов [1], който е скрил името на своята любима - поетесата Анна Ахматова, в началните букви на редовете:

Ангел лег у края небосклона.
Наклонившись, удивлялся безднам.
Новый мир был синим и беззвездным.
Ад молчал, не слышалось ни стона.
Алой крови робкое биение,
Хрупких рук испуг и содроганье.
Миру лав досталось в обладанье
Ангела святое отраженье.
Тесно в мире! Пусть живет, мечтая
О любви, о грусти и о тени,
В сумраке предвечном открывая
Азбуку своих же откровений

- **Метод на текстовите абревиатури (акроними)**

Друг метод за скриване на информация е използването на съкращения или акроними. M. Sirali-Shahreza и M. Hassan Shirali-Shahreza от Иран са предложили използването на заместване на думи с техните съкращения [8,11.]. Предложеният от тях метод работи, както следва:

Таблица 3

Акроним (0)	Превод (1)
218	Too late
2day	Today
ASAP	As Soon As Possible
C	See
U	You
CM	Call Me
F2F	Face to face

Текстови абревиатури (акроними)

Таблицата се състои от две колони, предварително организирана с избран списък от думи и съответните им акроними по такъв начин, че колоната с превода на акронима е "1", а съответното съкращение е с етикет "0". Този метод може да

има широко приложение в съвременните системи за текстова комуникация в реално време (чат системи). Широкото разпространение на тези системи, както и въведените като норма на общуване текстови абревиатури, позволяват предаване на стеганографска информация без никакви подозрения.

Емотиконите са емоционални икони, които се използват при онлайн чат. Тези емотикони изразяват чувството или настроението на хората, общуващи помежду си. Те могат да се разгледат и като абревиатури на по-дълги текстове, свързани с емоционални състояния. Използването на емотикони в стеганографията е доста интересно. Ванг, Чанг, Кию, Ли предлагат техника за текстова стеганография, базирана на емотикони [13].

- **Метод на семаграми**

Процесът на скриване на информацията чрез използването на знаци или символи се нарича семаграми. Тази техника включва картина, музика, чертеж, надпис или друг символ, за да се скрие информацията. Скриването на съобщение се осъществява и чрез промяна на външния вид на текст, като тип или размер на шрифта, добавяне на допълнителни интервали в него или разнообразни цветове. Текстовите семаграми се поделят на два вида - визуални и текстови.

Визуалните използват различно невинно изглеждащи обикновени символи и знаци, като драскулки, или позициониран обект в текста.

Текстовите семаграми скриват съобщението чрез промяна на външния вид на текста, чрез едва забележими промени в размера на шрифта или типа му, добавяне на допълнителни пространства (увеличаване или намаляване на разстоянието между буквите в текста, допълнителни интервали).

3.2. Техническа стеганография

Техническата стеганография използва физическото форматиране на текст като място, където да скрие информация. Обикновено този метод променя съществуващия текст, за да скрие стеганографския текст чрез вмъкване на интервали, умишлено написани правописни грешки разпределени в целия текст, оразмеряване на шрифта и т.н. Въпреки, че малко количество данни могат да бъдат скрити в документ, този метод може да се прилага за почти всички видове текст, без да разкрива съществуването на скрити данни. Компютърът може да не разпознава преоразмеряването на шрифта като проблем, особено ако се концентрира само върху текстово съдържание в документа, обаче човек може да открие странни размери на шрифта почти веднага. Освен това, ако е наличен оригиналният текст, сравнението на този текст с предполагаемия стеганографски текст би направило променените части от текста видими.

- **Метод на нулевия код**

Тук съобщението е скрито предварително според някой определени правила, например трябва да се чете всяка втора буква от първия, третия, петия ред и т.н.

През Втората световна война е използван и така наречения „нулев шифър" (т.е. некриптирано съобщение). Скриването на съобщение в огромен брой безполезни данни е нулев шифър. Той има вид на невинно съобщение, при което по предварителна уговорка - втората буква от всички думи, всяка пета дума и т.н. формират скрито послание.

- **Метод на произволните символи**

При този метод се генерира произволен низ, който съдържа единични букви като основния текст. Английските букви се разделят на две групи въз основа на тяхната форма, т.е. дали един символ има извивка във формата си или не (Таблица 4), дали един символ има една вертикална линия или не и т.н.

Впоследствие, когато искаме да скрием 0 бита във входния текстов файл, използваме буквите от група А сред генерираните букви, а когато искаме да скрием 1 бит, използваме буквите от група Б сред генерираните букви. [14].

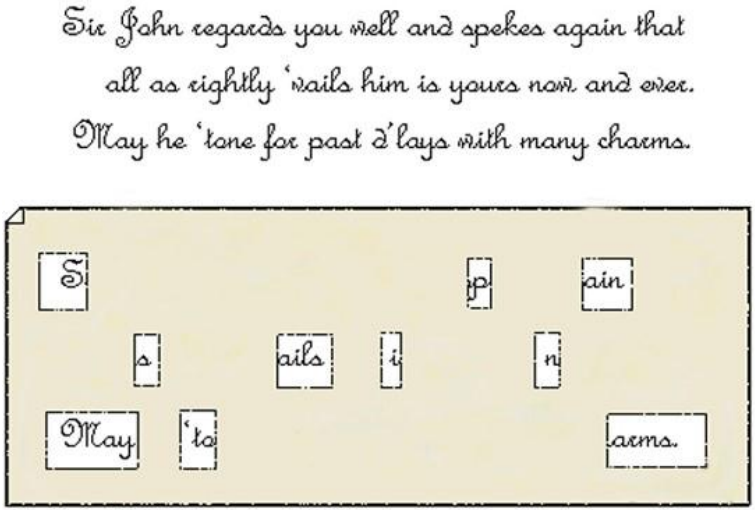
Таблица 4

Група	Име на групата	Bit	Букви
А	С извивки	0	B, C, D, G, J, O, P, Q, R, S, U
Б	Без извивки	1	A, E, F, H, I, K, L, M, N, T, V, W, X,Y, Z

Групиране на символите според извивката

- **Метод на скарите**

Този метод е известен още като е метода „Скара на Кардан“ (Фиг. 4) [1,15]. Тук се изписва текста като след това върху него се налага решетка с дупчици, през които се виждат определени букви. При прочита на тези букви се оформя „тайното“ съобщение.



Фиг. 4. „Скара на Кардан“

В новата ни история „скарите" са използвани от революционерите от периода на националното ни възраждане, в това число и от Апостола на свободата Васил Левски.

- **Метод на бели полета**

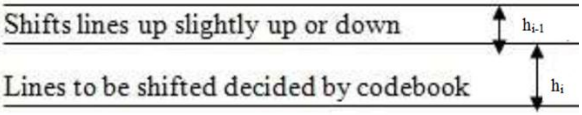
При този метод бялото пространство служи като основа за скриване на информацията [16]. Методът може да се използва по два различни начина: кодиране на линейно изместване и кодиране с преместване на думи.

Кодирането на линейно изместване е метод за промяна на документ чрез вертикално изместване на местоположенията на текстови редове, а кодирането чрез промяна на разстоянието между думите е метод за промяна на документ чрез хоризонтално изместване на местоположенията на думите в текстовите редове. [17]

Метод на линейното изместване

Тази техника променя документа, като вертикално измества позицията на местоположенията на текстовите редове [18,19]. Кодираната дума, предназначена за определен документ, определя текстовите редове, които ще бъдат преместени в този документ. Може да използваме „0" за линия (ред), изместена нагоре, и „1" за линия, изместена надолу. Енкодерът трябва да премества редовете нагоре или надолу, а декодерът измерва разстоянието между всяка двойка от два съседни реда. Това може да се направи с помощта на две различни техники: или декодерът измерва разстоянието между базовата линия на съседните редове, или декодерът измерва разстоянието между центроидите на два съседни реда. Базова линия е логическа линия, върху която са подредени символите на ред; центроидът е центъра на определен текстов ред. Да предположим, че текстовите редове i-1 и i+1 не са изместени и ред i е изместен или нагоре, или надолу. В непроменен текстов документ разстоянието между изходните редове е постоянно. Нека h_{i-1} и h_i са разстоянията между

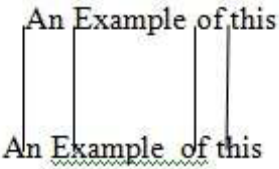
изходните редове i-1 и i и съответно между базовите линии i и i+1. Централното разстояние може да не е непременно равномерно разположено. При методи, които измерват разстоянието между центроидите, решението се основава на разликата между центроидните разстояния в оригиналния документ и в променения документ.



Фиг. 5. Метод на линейното изместване

Метод за скриване в разстоянието между думите

В този метод чрез хоризонтално изместване на думите и чрез промяна на разстоянието на думите, информацията се скрива в текста. Този метод е приемлив за текстове, където разстоянието между думите варира [12].



Фиг. 6. Метод на линейното изместване

Поради променливото разстояние, декодерът се нуждае от оригиналния документ или данни за разстоянието на думи в оригиналния документ. Първо кодиращият определя дали даден ред има достатъчен брой думи за кодиране, т.е. късите редове не са кодирани. На всеки намерен текстов ред за кодиране се прилага техниката на диференциално кодиране за тази схема. Втората, четвъртата, шестата и т.н. дума от лявото поле се измества. Първата и последната дума на всеки ред не са

изместени, за да се поддържа двустранното подравняване на колоната. След приключване на процеса на преместване на думи, документът се разпространява. Декодерът се нуждае от информация за оригиналния документ. Това не е недостатък, знаейки факта, че като цяло авторите проследяват документите си и притежават копие на оригиналния документ. Необходимата информация е позицията на началото на всяка дума или позицията на центроидите за всяка дума [20].

- **Промяна на характеристиките на текста**

В метода можем да променим характеристиките на текста един или повече пъти и по този начин променената функция може да послужи за основа на стеганографията. Функцията може да бъде стила, формата, цвета и размера на текста. Както например размерът на точката, използвана в малките английски азбуки i и j, може да бъде променена, за да се скрие 0 или 1. Такива характеристики могат да бъдат вертикалните линии на буквите b, d, h, k и др. Дължината на тези линии може да се променя по начин, който е незабележим с просто око. Височините на символите в рамките на даден шрифт също могат да бъдат променени. Може да се използва за вграждане на информация за авторските права, а не само за криене на информация.

Една от характеристиките на тези езици е изобилието от точки в буквите му. Буквите с една точка могат да бъдат използвани за скриване на информацията, като се измести позицията на точката малко вертикално по отношение на стандартната й позиция в текста.

3.3. Стеганография в програмни кодове

Езиците за програмиране се подчиняват на същите синтактични и семантични правила като естествените езици. С тяхна помощ се създават програмни кодове, чиято структура е подобна на тази на обикновените текстови файлове.

При повечето езици от високо ниво се прави компилиране на програмния код, в резултат на което се получава нов тип файл, наречен изпълним код. В тези случаи, вграждането на стеганографска информация в текста на програмния код, ще доведе до невъзможност за нейното възстановяване.

При част от програмните езици не е необходима компилация на програмния код, а се прави директна интерпретация на кода от програмна среда или приложение. Типичен представител на тези езици са така наречените „markup“ езици. Те са особено благоприятни за вграждане на скрита информация.

- **Метод на бели полета**

В този метод данните се вграждат чрез вмъквани на бели полета в HTML таговете. При този метод се добавя допълнително място в съответствие с тагове. Полето се вмъква след четене на символа "<" и преди четене на ">". Белите полета представляват еднобитови данни в HTML файлове. Обратна процедура се прилага във фаза на извличане, когато всички допълнителни интервали в тага са премахнати [21].

- **Метод на замяна на HTML таговете**

При този метод първо се избира подходящия HTML файл. Търси се в кой DIV таг да се скрие тайното послание и след това данните се вграждат чрез добавяне на параметър вътре в DIV елемента [7].

HTML таговете може да се използват в различни комбинации [12]:

Стего ключ
<img></img> -> 0
<img/> -> 1

Стего файл
<img src=g1.jpg></img>

<img src=g5.jpg></img>

Скрити битове: 01110

Тук данните се вграждат с помощта на празни тагове. Представянето на празен таг е или начален таг, последван от таг за край, или празен. Обикновено тази техника може да бъде използвана или приложена с помощта на <img> тага. В този метод първия таг на изображението се взема и затварящия символ „/” се добавя преди прочитане на символа за край „>”. За да приключи процеса на извличане трябва да имаме и двете <tag/> и <tag></tag>. Така че, когато символа „/” се изтрие от първия таг <tag/>, то ще има друг затварящ таг </tag>, за да се избегне евентуална грешка. В обратния процес „/” се изтрива преди четене на знака за край „>” [22].

В този метод се създават и скрити таблици. Избират се подходящото послание, което ще бъдат добавено в таблицата. Това става като всяко послание се добавя в ред от таблицата [7].

Среща се и промяна на малки и главни букви (регистъра) в HTML таговете. Те са нечувствителни към регистъра, поради което можем да се възползваме от него, за да скрием съобщение в документа, като променим регистъра на конкретни букви в името на тага Например, <ID>, <id>, <Id> и <iD> означават абсолютно едно и също и можем да кодираме два бита, като изберем една от неговата версия. Големият капацитет е основното предимство на този метод. От друга страна е много лесно да се открие стего каналът, тъй като е много необичайно да се използват редуващи малки и главни букви [23].

- **Метод на разделяне на думите**

В метода на разделяне на думите данните се вграждат чрез тага <p>. В този метод текста първо се разделя на блокове от думи. След това битовете се вграждат чрез коригиране на ширината на разстоянията между символите в рамките на един блок, в предварително зададено правило. Размерите на блока от 18-20 символа се предефинират и се вгражда 3 бита информация [7].

- **Метод на скриване в атрибутите**

Този метод скрива тайната информация чрез използване на HTML тагове и атрибути. Една възможност е скриване на съобщенията чрез промяна на реда на атрибутите, тъй като подреждането на атрибутите не влияе на външния вид на HTML документите, които са основни елементи на мрежата. Тези документи се използват много често в Интернет и следователно са по-малко склонни да предизвикат подозрение за съществуването на тайното съобщение. Освен това, всеки HTML документ има значителен брой тагове и атрибути. По този начин капацитетът на процеса на скриване на секретни съобщения също е висок в предлаганата техника.

Някои от HTML атрибутите имат дефинирани стойности по подразбиране. HTML документ се възприема по същия начин, без значение дали стойностите по подразбиране са изрично дефинирани или не. Това дава възможност да се скрие допълнителна информация чрез посочване на стойности по подразбиране в някои части на HTML документа и пропускането им в други части. Този метод е труден за откриване, но ограниченият възможен брой атрибути със стойности по подразбиране е основния му недостатък.

HTML стандартът не определя реда от атрибутите, което означава, че всеки ред може да се използва, без да се засяга външния вид на уеб страницата. Тъй като редът на атрибутите няма значение, този метод може да се прилага без ограничения. Промяната на реда на атрибутите за скриване на информация в

HTML документ е най-интересният метод за HTML стеганографията. Това не променя оригиналния размер на файла и е трудно да се открие без компютърни програми, анализиращи структурата на HTML документа.

Имайки таг, включващ 8 атрибута, има 8! = 40320 различни пермутации, което позволява да се скрие над 15 бита информация в рамките на един и същ таг. Този метод вероятно е най-често споменаваният в контекста на скриване на данни в документите за markup езика. Основното му предимство се крие в неговата сигурност, но на практика позволява да се изпраща само малко количество данни, тъй като е ограничено от броя на атрибутите, използвани в оригиналния документ.

- **Метод на произволните символи**

В този метод данните се вграждат чрез вмъкване на произволни знаци в таговете. Символ се вмъква след прочитане на първия символ от първия таг. По същия начин след всяка дума се вмъква по един случаен символ. В случай на специален символ, процесът се повтаря отново. Процесът на вграждане се прилага рекурсивно към всички тагове. При обратния процес всички вмъкнати символи се изтриват от файла.

- **Метод на смяната на цвета**

При метод на смяната на цвета данните се вграждат чрез замяна на името на цвета с неговата шестнадесетична стойност. В този метод първо трябва да намери атрибута на цвета, последван от символ "=" и след това името на цвета се заменя с неговата шестнадесетична стойност. В обратен ред шестнадесетичната стойност се връща обратно към нейното име [7].

4 Заключение

Възможност за комбинации между грешки, знаци, разстояния между символите, влагане на „жаргони” (условности) и т. н. правят текстовата стеганография неоткриваема и на

практика най-сигурна и неподдаваща се на анализи. За това способства и обстоятелството, че не е нужна математическа логика, което често прави неефективни компютърните системи и програмните продукти за анализ на съобщенията. Използването на такъв тип стеганографски методи позволява лесна програмна реализация на вграждане и извличане на стеганографска информация. Това прави толкова привлекателно използването им при обучение в областта на защита на информацията и информационна сигурност.

ЛИТЕРАТУРА:

[1] Станев, С. Стеганологична защита на информацията, Университетско издателство „Епископ Константин Преславски". Шумен, 2013. ISBN 978-954-577-825-4. 320.

[2] Станев, С., Железов, С., Параскевов, Х., Христов, Х., Ръководство за упражнения по стеганография, Университетско издателство, Шумен, 2015, ISBN 978-619-201-011-9.

[3] Stanev, S., Szczypiorski, K. Steganography Training: A Case Study from Univeristy of Shumen in Bulgaria. *International Journal of Electronics and Telecommunications,* 2016, vol. **62**, no. 3, PP. 315-318

[4] Станев, С., В. Галяев.Смысловое сопостав ление научных терминов на русском и английском языках в области компьютерной стеганографии, *Сборник материалов международной научно-практической конференции*. ГАОУ ВПО "Дагестанский государственный институт народного хозяйства". – Махачкала.: ДГИНХ, 2013. стр.51-56.

[5] Параскевов, Хр., Стефанов, Ал., Съвременни стеганографски подходи в социалните мрежи, MATTEX 2018, Том **1**, стр. 197-203.

[6] Kordov, K., Stoyanov, B. (2017). Least Significant Bit Steganography using Hitzl-Zele Chaotic Map. *International Journal of Electronics and Telecommunications*, Vol. **63**, No. 4, pp. 417-422

[7] Agarwal, Monika, Text steganographic approaches, *International Journal of Network Security & Its Applications* (IJNSA), Vol. **5**, No.1, 2013, pp 91-106.

[8] M. H. S. Shahreza, and M. S. Shahreza, A new approach to Persian/Arabic text steganography. *Proceedings of 5th IEEE/ACIS Int. Conf. on Computer and Information Science and 1st IEEE/ACIS Int. Workshop on Component-Based Software Engineering, Software Architecture and Reuse*, 2006, pp. 310-315.

[9] Anderson, R., F. Petitcolas. On The Limits of Steganography. *IEEE Journal of Selected Areas in Communication*, 1998. Special Issue of Copiright & Privacy Protection. ISSN 0733-8716, **16**(4):474-481

[10] M. H. Shirali-Shahreza, M. Shirali-Shahreza, A new approach to persian/arabic text steganography, *Proc. 5th Int. Conf. Computer andInformation Science*, Washington, 2006, pp. 310-315.

[11] M. Shirali-Shahreza, Text Steganography by Changing Words Spelling, In: 10th International Conference on Advanced Communication Technology, 2008, ICACT 08, vol. **3**, pp. 1912-1913.

[12] Prem Singh, Rajat Chaudhary and Ambika Agarwal, A Novel Approach of Text Steganography based on null spaces, *IOSR Journal of Computer Engineering* (IOSRJCE), 2012, Volume **3**, Issue 4, pp 11-17.

[13] Z.H. Wang, C.C. Chang, D. Kieu, and M.C. Li, Emoticon-based Text Steganography in Chat, Second Asia-Pacific Conference on Computational Intelligence and Industrial applications, 2009, ISBN: 978-1-4244-4606-3.

[14] Shraddha Dulera, Devesh Jinwala and Aroop Dasgupta, Experimenting with the novel approaches in text steganography, *International Journal of Network Security & Its Applications (IJNSA)*, Vol. **3**, No.6, November 2011, pp 213-220.

[15] Апостолов, Д. и С. Станев, Софтуерна реализация на класическия стеганографски способ "Gardan grill". *Годишник на Факултета по технически науки*, ШУ, 2013.

[16] Ray, Rishav, Jeeyan Sanyal, Debanjan Das, and Asoke Nath, A New Challenge of Hiding any Encrypted Secret Message inside any

Text/ASCII File or in MS Word File: RJDA Algorithm, International Conference on Communication Systems and Network Technologies, May 2012, vol. – **6**, pp. 889-893.

[17] L. Y. Por, T. F. Ang and B. Delina, WhiteSteg: A new scheme in information hiding using text steganography, *WSEAS Transactions on computers*, 2008, Issue **6**, Volume **7**, pp 735-745.

[18] S.H. Low, N.F. Maxemchuk, J.T. Brassil, and L. O'Gorman, Document marking and identification Using both line and word shifting, *Proceedings of the Fourteenth Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM '95)*, 2-6 April 1995, vol. **2**, pp. 853 - 860.

[19] Richard Popa, An Analysis of Steganographic Techniques, The Politehnica University of Timisoara, Faculty of Automatics and Computers, Department of computer science and Software Engineering. 1998.

[20] Hitesh Singh, Pradeep Kumar Singh, Kriti Saroha, A Survey on Text Based Steganography, *Proceedings of the 3rd National Conference; INDIACom-2009*, Computing For Nation Development, February 26 – 27, 2009

[21] Dhammjyoti V. Dhawase, Sachin Chavan, Webpage information hiding using page contents, *IJARCET*, Volume **3**, Issue 1, January 2014, pp 182-186.

[22] Zhelezov, S., Uzunova-Dimitrova, B., Paraskevov, H., An approach for hiding steganography data within web applications, *Journal of Engineering and Applied Sciences* **12**(Special issue 8), pp. 8251-8255.

[23] L. Polak1, Z. Kotulski, Sending hidden data through WWW pages: detection and prevention, 2010 Polish Academy of Sciences Institute of Fundamental Technological Research, **58**, 1–2, 75–89.

Теодора Тихомирова Стоянова
ШУ„Еп. Константин Преславски“
E-mail: t.stoyanova@shu.bg

Станимир Кунчев Железов
ШУ„Еп. Константин Преславски“
E-mail: s.zhelezov@shu.bg