Digitalna forenzika

Tekstualna steganografija - Metode sakrivanja teksta korišćenjem praznog prostora i nevidljivih karaktera u Word dokumentima

Mentor:

Prof. dr Bratislav Predić

Student:

Natalija Mitić, 1046

Sadržaj

Uvod

Steganografija je tehnika sakrivanja tajnih podataka unutar obične, netajne, datoteke ili poruke kako bi se izbeglo otkrivanje. Tajni podaci se zatim izvlače na odredištu. Steganografija se može koristiti za prikrivanje skoro bilo koje vrste digitalnog sadržaja, uključujući tekst, sliku, video ili audio sadržaj. Takođe, podaci koje treba sakriti mogu se sakriti unutar gotovo bilo koje druge vrste digitalnog sadržaja.

U ovom radu je obrađena tekstualna steganografija, odnosno prikazane su metode skrivanja teksta u *word* faju. Tekstualna steganografija predstavlja izazov, jer tekst ne sadrži puno redundantnih bitova koji se mogu koristiti za skrivanje tajnih podataka, kao što je to slučaj kod slika ili audio i video fajlova. Zbog toga je kapacitet ugradnje uglavnom manji, a što veći kapacitet je jedna od bitnijih karakteristika kojima algoritmi teže.

Cilj rada je da prikaže metode koje se zasnivaju na praznim prostorima (razmacima između reči), kao i karakterima koji nemaju pisani trag. Pošto se sadržaj koji se skriva steganografijom, često se šifrira pre nego što se ugradi u datoteku nosioca kako bi se povećala sigurnost, obrađena je i metoda koja enkriptuje poruku. Takođe, razmotrena je i kompresija poruke prilikom skrivanja, kako bi se povećeo kapacitet ugradnje.

Steganografija

Steganografija je naučna disciplina koja se bavi prikrivenom razmenom informacija. Naziv je nastao kombinacijom dve grčke reči – *steganos* (skriveno) i *graphein* (pisanje). Dakle, ovaj pojam se odnosi na skriveno pisanje, odnosno prikrivanje neke tajne poruke.

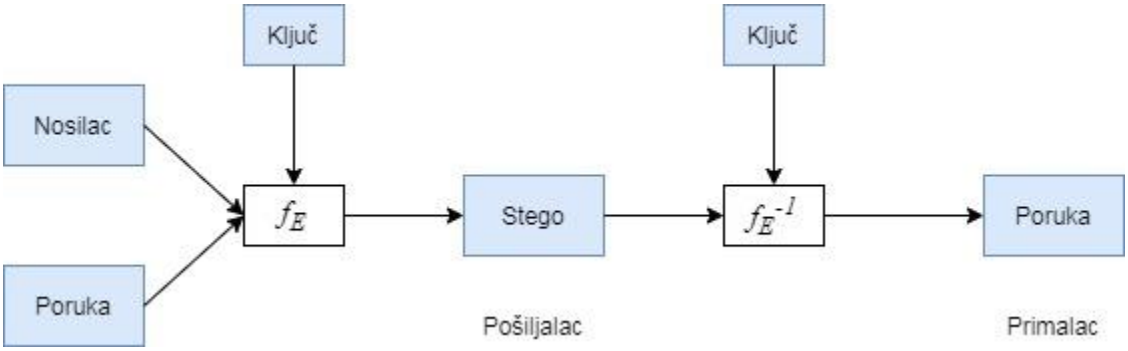
Steganografija datira još od 200. godine pre Hrista, iako ne u obliku u kakvom je danas poznata. Jedan od takvih primera su Naska linije u Peruu[1]. Na suvoj ravni je izgrebano više od 80 km linije između dva grada Naska i Palpe, od kojih su mnoge vidljive tek iz vazduha.

Savremeni pojam steganografije, odnosno steganografskog sistema (stegosistema), podrazumeva skup sredstava i metoda koji se koriste za formiranje skrivenog kanala prenosa informacija.

Proces steganografije obično uključuje ugrađivanje tajne poruke unutar nekog prenosnog medija. Opšti proces, prikazan na slici 1, dat je relacijom:

steganografski_medij = tajna_poruka + nosilac poruke + steganografski_ključ

[1] https://en.wikipedia.org/wiki/Nazca_Lines



Slika 1 – Opšti proces steganografije

U svojstvu podataka može se koristiti bilo koja informacija, odnosno poruka (tekst, audio podaci ili slika). Noslilac poruke (eng. carrier, cover, cover medium) je bilo koja informacija namenjena da kao nosilac prenese skrivenu poruku. Ugrađena ili tajna poruka, (eng. embedded message) je poruka koja se implementira u nosioca poruke. Stego ključ (eng. stego-key) je tajni ključ pomoću koga se tajna poruka implementira u nosioca poruke. U stegosistemu može postojati jedan ili više stego ključeva, a po analogiji sa kriptografijom, razlikujemo stego sisteme sa tajnim i javnim ključem. Stego ključ je parametar funkcije f_E koja služi za „ugrađivanje“ poruke u nosioca. Steganografski medijum (eng. steganography medium, stego-medium) je posrednik koji sadrži implementiranu poruku koja se tajno prenosi. Steganografski kanal (stegokanal) je komunikacioni kanal preko kojeg se šalje stego nosilac poruke. Funkcija f_E^{-1} služi za „izdvajanje“ tajne poruke.

Primena steganografije se najčešće bazira na sledećem principu: pošiljalac tajne poruke bira nasumično nosioca poruke; u izabrani nosilac poruke implementira se tajna poruka uz pomoć stego ključa; primaocu se šalje steganografski medijum, a primaoc na drugoj strani obrnutim postupkom dolazi do sadržaja tajne poruke. Da tajna poruka ne bi bila vidljiva, bitno je da nosilac poruke sadrži dovoljno redundantnih bitova, koji mogu biti zamanjeni tajnom porukom. Važno je napomenuti da nisu svi digitalni formati pogodni za prenošenje tajnih poruka, jer se npr. promenom bitova u nekom izvršnom fajlu dovodi do toga da program ne radi ili javlja greške pri radu.

Steganografija ima vrlo široke mogućnosti primene - od prikrivene razmene podataka u privatne i poslovne svrhe, pa sve do zaštite autorskih prava u obliku vodenog žiga. Upotrebom seganografije može se obezbediti poverljivost važnih informacija, čuvajući ih od sabotaže, krađe ili neovlašćenog gledanja. Međutim, zbog svog principa „nevidljivosti” informacija, često se koristi i za razne ilegalne aktivnosti. Njihovo svakodnevno korišćenje ove tehnike zadalo je velike muke, kako stručnjacima zaduženim za sigurnost, tako i kompjuterskim forenzičarima, ali i sudskim veštacima prilikom istraga kriminalnih dela.

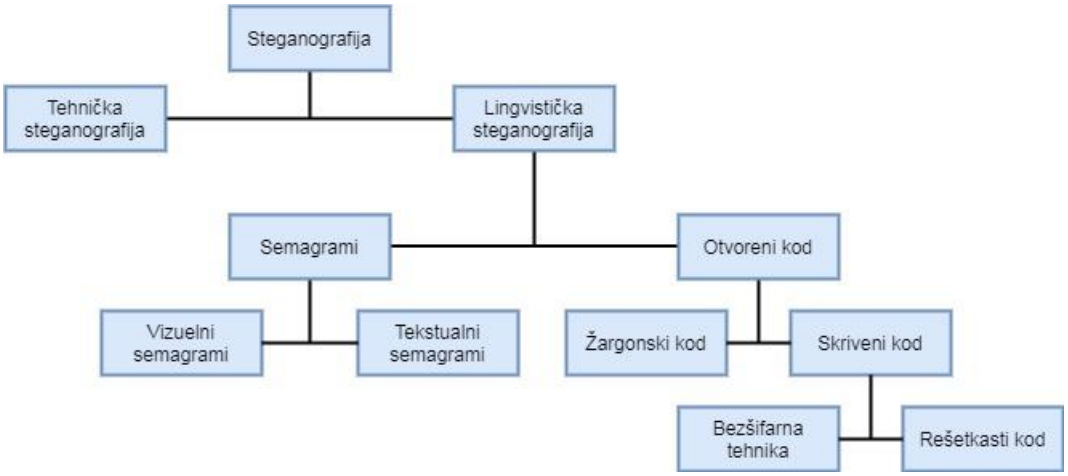
Za zaštitu informacija najčešče se koriste metode kao što su kriptografija, kodiranje ali i steganografija. Osnovna prednost steganografije u odnosu na kriptografiju je činjenica da poruke ne privlače pažnju na sebe. Može se reći da je steganografijom moguće izbeći napad „čovek u sredini“ (eng. man in the middle), s obzirom na to da napadač nije svestan postojanja komunikacije u nekom komunikacionom kanalu, dok je kod kriptografije vidljivo da strane komuniciraju kroz kriptovan kanal. Steganografija ima prednost u zemljama u kojima su kriptografske tehnike za šifrovanje poruka zabranjene.

Klasifikacija steganografskih tehnika

Tehnike steganografije se mogu podeliti u više grupa:

- Tehnička steganografija koristi naučne metode za sakrivanje poruka, kao što je upotreba nevidljivog mastila ili mikrotačaka i druge metode smanjenja veličine tajnih poruka.
- Lingvistička steganografija krije poruke u nosioce na neki neuobičajan način i dalje ih kategoriše kao semagrame ili otvorene kodove.
- Semagrami kriju informacije kroz simbole ili znakove. Vizuelni semagrami koriste naizgled bezopasne ili svakodnevne fizičke oblike za prenos poruka, kao što su postavljanje detalja na desktop ili web sajt. Tekstualni semagrami kriju poruke modifikujući izgled nosioca teksta, kao što je vešta promena tipa ili veličine slova, dodavanjem dodatnih praznih mesta, ulepšavanjem slova ili dodavanje ručno pisanog teksta.
- Otvoren kod krije poruke u legitimne nosioce poruka na način koji nije očit nekom nesumnjičavom posmatraču. Nosilac poruke se ponekad zove javno saopštenje, s obzirom na to da je skrivena poruka prikriveno saopštenje. Ova kategorija je podeljena na žargonski kod i sakriveni kod.
- Žargonski kod, kao što mu ime kaže, koristi jezik koji je razumljiv grupi ljudi, ali je beznačajan drugima. Žargonski kod uključuje simbole koji se koriste da ukažu na prisustvo i vrstu signala bežične mreže, podzemnu terminologiju, ili bezazlenu konverzaciju koja prenosi posebno značenje zato što su činjenice poznate samo govorniku. Podskup žargonskog koda je i znakovni kod u kome se, izvesnim preuređenjem fraze, menja značenje.
- Sakriveni ili tajni kod krije poruku javno u nosećoj poruci tako da ona može biti otkrivena od bilo koga ko zna na koji način je ona sakrivena. Rešetkasti kod koristi šablon koji je korišćen za prikrivanje u nosiocu poruke. Reči koje se pojave pri otvaranju šablona su sakrivena poruka. Bezšifarna tehnika krije poruku prema nekom preuređenom setu pravila, poput „čitaj svaku petu reč” ili „traži svako treće slovo u svakoj reči”.

Na slici 2 je grafički predstavljena podela steganografskih tehnika.



Slika 2 – Klasifikacija steganografskih tehnika

Kompjuterska steganografija

Da bi se sakrila poruka u digitlnom fajlu koristeći bezšifarnu tehniku nisu potrebni specijalni alati ili posebne veštine. Slika ili tekstualni blok može biti sakriven u drugoj slici u PowerPoint fajlu, na primer. Poruka može biti sakrivena u atributima Word fajla, u komentaru na Web stranici ili u nekom drugom formatu kojeg veb pretraživač ignoriše. Tekst može biti sakriven kao stilizovana linija u dokumentu tako što se tekstu dodeljuje boja pozadine, a zatim se on postavlja u drugi crtež koji je u prvom planu. Primalac može da povrati skriveni tekst tako što će mu promeniti boju. Ovo su sve svakako nisko-tehnički mehanizmi, ali su veoma efektni.

Pomoću steganografije informacije je moguće skriti unutar:

- Slike-fotografije (.bmp, .gif, .jpeg i sl.)
- Video fajla (.avi, .mpg, .vob i sl.)
- Audio fajla (.mp3, .midi, .wav, .wma i sl.)
- Datoteke (.doc, .xls, .ppt, .txt i sl.)
- Bilo kojeg binarnog fajla

Savremeni metodi kompjuterske steganografije često se zasnivaju na statističkom preobilju informacija u audio i video digitalnim signalima. Preobilje informacija u audio i video digitalnim signalima je osnovni pravac razvoja metoda kompjuterske steganografije. Digitalna fotografija, digitalna muzika i digitalni video predstavljaju matricu brojeva koji kodiraju intenzitet (jačine svetlosti kod fotografije, intenzitet jačine zvuka kod muzike) u diskretnom trenutku u vremenu i prostoru. Kako bitovi sa najmanjim značenjem sadrže malo korisnih informacija, to njihova zamena tajnom porukom ne utiče značajno na kvalitet reprodukovanog zvuka ili slike. Kako zbog grešaka kvantizacije, digitalne signale već prati šum, zvuk se neznatno pogoršava i ne može se registrovati čovekovim čulima.

Steganografija se može podeliti na 2 vrste u pogledu robusnosti: *Fragile* (lomljiva, krhka) i *Robust* (robusna, snažna). Pomoću fragilne steganografije informacija se umeće u fajl pri čemu, ukoliko dođe do promena na fajlu nosiocu, dolazi do potpunog gubitka informacije. Kod robusne steganografije, informacija se umeće u fajl pri čemu je veoma teško ovakvu informaciju uništiti ili oštetiti. Ovaj tip steganografije je mnogo teži za implementaciju, ali su mu i mogućnosti primene mnogo veće.

U steganografiji postoji nekoliko različitih tehnika koje se mogu koristiti za skrivanje informacija u datotekama. To su:

- Ubacivanje (eng. Injection, Insertion)
- Zamena
- Generisanje

Ubacivanje

Ova tehnika omogućava skrivanje postojanja podataka u delovima datoteka koji su od manjeg značaja za zlonamernog korisnika. Tehnika se bazira na dodavanju bitova u datoteke tako da površinski deo datoteke ostane savršeno čist. Dodavanje određenog broja dodatnih

bezopasnih bitova u izvršnu datoteku neće bitno uticati na proces koji se izvršava, a prisustvo metode neće se odraziti na konačan ishod metode, tako da krajnji korisnik ne može osetiti prisustvo skrivenog podatka u datoteci. Međutim, upotreba tehnike umetanja menja veličinu datoteke u zavisnosti od ukupnog broja utisnutih bitova, što može dovesti da neuobičejeno velika datoteka izazove određenu sumnju kod zlonamernog korisnika.

Zamena

Pristup zamene zasniva se na zameni najmanje značajnih bitova datoteke, i to na takav način da primena ove metode ima što manji efekat na izobličenje originalne datoteke. Glavna prednost ove tehnike je u tome što se veličina datoteke ne menja prilikom primene steganografskog algoritma. S druge strane, ova metoda ima i dva nedostatka. Prvi je degradacija steganografski obrađene datoteke i ograničenje broja manje značajnih bitova koji se mogu upotrebiti za primenu ove metode.

Generisanje

Nedostatak kod prethodno pomenute dve tehnike - ubacivanja i zamene, je taj što se originalna datoteka može porediti sa stego datotekom i tom prilikom moguće je otkriti razlike. Tehnika generisanja ne zahteva originalnog nosioca podataka, već sama generiše datoteku u kojoj će biti sadržana poruka. Kada se koristi tehnika generisanja, konačan rezultat je originalna datoteka koja je imuna na komparaciju sa drugim datotekama.

Steganografske metode

U suštini, steganografija koristi ograničene sposobnosti našeg vizuelnog sistema. Bilo koji otvoreni ili šifrovani tekst, slika i sl., korišćenjem određenih metoda može biti utisnut u određenu sliku nosioca podataka, a da to ne bude vidljivo okom. Postoji više metoda koje se koriste za skrivanje informacija unutar teksta, slike, audio ili video datoteke. Neke od metoda koje se koriste u steganografiji su:

- HTML stego
- LSB (Least Significant Bit)
- SS (Spread Spectrum)
- DCT (Diskretna Kosinusna Transformacija)
- WS (White Space) metoda

HTML stego metoda

Ova metoda koristi tehniku zamene. Bazira se na skrivanju informacija u izvornom HTML kodu tako da podaci koji se prezentuju korisniku ostaju nepromenjeni. Tehnika se zasniva na zameni bitova manje važnih identifikatora decimalnog oblika koji se odnosi na boju teksta sa njihovim ekvivalentima u tekstualnom obliku. HTML metoda se najviše koristi za obmanu računara koji u potrazi za informacijama pretražuju internet.

LSB metoda

LSB (Least Significant Byte), ili bit najmanje težine, je metoda koja koristi tehniku zamene vrednosti najmanje značajnih piksela u binarnom obliku. U datoteci se obično nalazi nekoliko bitova koji nisu stvarno potrebni ili na određeni način nisu toliko važni. Ti bitovi u datotekama mogu poslužiti za prenos skrivenih informacija na taj način da neće bitno menjati datoteku ili je oštetiti. LSB metoda je najbolju primenu našla u slikovnim datotekama koje imaju visoku rezoluciju uz upotrebu različitih boja i u audio datotekama koje reprodukuju različite zvukove na velikim brzinama. LSB metoda obično ne povećava veličinu datoteke, ali zavisno od veličine informacije koja se skriva, može primetno deformisati datoteku.

Spread Spectrum metoda

Ova metoda koristi tehniku ubacivanja a zasniva se na širenju frekvencijskog spektra signala u određenom domenu. Dodaju se šumovi u slučajno odabrane signale. Metoda koristi slabosti koje imaju ljudski organi čula. Takođe, koristi se i za kontrolu bezbednosti komunikacionog kanala, povećanje otpornosti na prirodne smetnje, sprečavanje otkrivanja i za ograničenje snage određenih prenosnih linkova. U audio steganografiji implementacija je moguća pažljivim biranjem audio sadržaja u koji se utiskuju podaci. Trenutne steganografske aplikacije koje koriste ovu metodu su, pre svega, ograničene na potvrdu dokaza o autorskim pravima, kao i garancijama integriteta sadržaja. Pošto se koristi tehnika zamene niskih talasa, slično kao kod LSB metode, problem kod ove metode je što su niski talasi uočljivi za ljudsko uho tako da je ovo praktično prilično ranjiva metoda.

Slična ovoj metodi je i *ECHO* metoda koja takođe koristi tehniku ubacivanja podataka. Ova metoda za skrivanje informacija koristi odjeke u audio datotekama. Princip rada se zasniva na jednostavnim ubacivanjem dodatnog zvuka eha unutar audio datoteka u kom je sadržana informacija. Ono što čini ovu metodu izdvaja od ostalih audio-steganografskih metoda je to da se zapravo audio zvuk unutar datoteke može dodatno poboljšati.

DCT metoda

Steganografske metode koje skrivaju podatke u bitove najmanje važnosti neke slike su proste i efikasne, ali krhke. Svaka modifikacija slike može da promeni osobine piksela i time uništi skrivene podatke. Zato steganografska metoda treba da bude snažna, odnosno treba da zadrži skrivene podatke i posle modifikacije piksela. Jedan od pristupa je da se prvo slika transformiše, a zatim da se u nju ugrade tajni podaci. Podaci se ugrađuju u transformisane piksele i slika se transformiše u širi domen. Slika se sada može modifikovati, i kada se takva slika ponovo transformiše, skriveni podaci će i dalje biti prisutni u modifikovanim pikselima.

Jedan način primene opisanog postupka je diskretna konusna transformacija (DCT). Ona konvertuje originalne piksele u brojeve (koeficijente transformacije) koji označavaju frekvenciju sadržanu u slici. Prostornu frekvenciju vrste slike definišemo kao broj promena boje u jednoj vrsti. DCT konvertuje blokove piksela u blokove DCT koeficijenata koji odgovaraju prostornoj frekvenciji slike. Podaci se mogu skrivati u blokovima DCT koeficijenata uz izračunavanje DCT transformacije slike i neznatno menjanje nekih od koeficijenata, posle čega se vrši inverzna DCT sa modifikovanim pikselima, da bi se slika

vratila u početno stanje. Modifikovani koeficijenti se ne smeju nalaziti blizu gornjeg levog, kao ni blizu donjeg desnog ugla, jer oni odgovaraju niskoj frekvenciji slike i smatraju se važnim, dok se ostali koeficijenti mogu značajno promeniti.

White Space metoda

White Space metoda je još jedna steganografska metoda koja se bazira na tehnici ubacivanja. Dodavanjem dodatnog praznog prostora unutar pisanog teksta na listovima papira jednostavnim pritiskom tastera *space* na tastaturi, običnom korisniku verovatno neće izazvati pažnju. Prazni prostori su uobičajena pojava u svim dokumentima koji se svakodnevno koriste, pa je primena ove metode veoma efikasna za većinu tekstualnih datoteka. Ova metoda se može primenjivati u skoro svim datotekama u kojima je smešten tekst za čitanje: doc, pdf, rtf.

Steganografija teksta

Skrivanje tajne poruke u tekstu

Tekst je jedan od najstarijih medija koji se koristi u steganografiji. Korišćen je mnogo pre elektronskog doba, kada su pisma, knjige i telegrami skrivali tajne poruke u svojim tekstovima.

Tekst predstavlja najteži fajl za skrivanje podataka, jer sadrži malo redundantnih podataka u poređenju sa slikama, audio i video fajlovima. S druge strane, neki steganografski algoritmi zavise od osobina jezika, pa se možda ne mogu implementirati u svim jezicima.

Struktura tekstualnog dokumenta je obično vrlo slična onome što se vidi, dok je u svim ostalim vrstama medija (audio, slika, video) struktura drugačija od one koju opažamo, čineći skrivanje informacija jednostavnim i bez uočljive izmene. Prednost steganografije koja skriva podatke u tekstu, u odnosu na druge medije, je manje zauzimanje memorijskog prostora, jednostavnija komunikacija, mogućnost slanja više informacija, a troškovi za štampanje su manji.

Danas su računarski sistemi olakšali sakrivanje informacija u tekstovima. Takođe se proširila i upotreba skrivanja podataka u tekstu. Među najvažnijim od ovih tehnologija može se navesti skrivanje informacija u elektronskim tekstovima, veb stranicama i dokumentima.

Pošto svi mogu da čitaju, kodiranje teksta u neutralnim rečenicama može delovati neefikasno. Međutim, uzimanjem prvog slova svake reči neke rečenice, može se zaključiti da je to moguće i da nije jako teško. Sakrivanje informacija u običnom tekstu može se izvršiti na mnogo različitih načina. Algoritam prvog slova koji je spomenut nije baš siguran, jer znanje o sistemu koji se koristi automatski odaje tajnu poruku. To je nedostatak koji mnoge tehnike skrivanja podataka unutar običnog teksta imaju kao zajedničko.

Mnoge tehnike uključuju modifikaciju izgleda teksta, pravila poput korišćenja svakog n-tog znaka ili promene količine praznog prostora posle redova ili između reči. Poslednja tehnika uspešno je korišćena u praksi, pa čak i nakon što je tekst štampan i kopiran na papir deset puta, tajna poruka je i dalje mogla da se preuzme.

Drugi mogući način čuvanja tajnih podataka unutar teksta je upotreba javno dostupnog izvora nosioca, knjige ili novina i upotrebe koda koji se sastoji, na primer, od kombinacije broja stranice, broja linije i broja znaka. Na ovaj način, nikakve informacije, sačuvane u nosiocu neće dovesti do skrivene poruke. Otkrivanje zavisi isključivo od sticanja znanja o tajnom ključu.

Postoje tri osnovne kategorije steganografije teksta: metode zasnovane na formatu, slučajna i statistička generacija i lingvističke metode. U svakoj od ovih kategorija tekst se može generirati ispočetka ili ugraditi u poznati tekst (plaintext).

Metode zasnovane na formatu

Metode zasnovane na formatu koriste fizičko formatiranje teksta kao prostora u kojem se mogu sakriti informacije. Metode zasnovane na formatu uglavnom menjaju postojeći tekst kako bi sakrile steganografski tekst. Umetanje razmaka ili neprikazanih znakova, namerno pravopisno pogrešno napisana slova kroz tekst i promena veličine fontova, neki su od mnogih metoda steganografije zasnovanih na formatu. Neke od ovih metoda, kao što su namerne pravopisne greške i umetanje praznog prostora, mogu zavarati ljude koji čitaju tekst i koji ignorišu povremene pravopisne greške, ali ih računar često može lako otkriti.

U ovoj metodi, poruka nosilac se neće menjati u pogledu njenih reči i rečenica. Modifikacije će se izvršiti samo na razmacima između reči, linija i / ili odlomaka pomoću posebnih znakova, odnosno pomoću steganografije praznog prostora. Pri korišćenju ove metode, broj dodatnih specijalnih znakova (npr. razmak) ostalim razmacima u poruci nosiocu zavisi od vrednosti cifara binarnog niza.

Ova metoda zavisi od osobina jezika, pa se ne mogu svi algoritmi implementirati u svim jezicima, kao što je to slučaj sa jezicima koji nemaju razmake između reči.

Metode slučajne i statističke generacije

Da bi se izbeglo poređenje sa poznatim tekstom, steganografi često pribegavaju stvaranju sopstvenih tekstova nosioca. Iako ovo često rešava problem napada na poznatog nosioca (eng. known cover attack), svojstva generisanog teksta i dalje mogu da izazovu sumnje da tekst nije legitiman. Takva generacija generalno pokušava da simulira neko svojstvo normalnog teksta, obično približavanjem neke proizvoljne statističke distribucije pronađene u stvarnom tekstu.

Jedan pristup steganografije teksta može sakriti informacije u nečemu što se čini slučajnim nizom znakova. Naravno, i osobi koja šalje i prima poruku, ova sekvenca je daleka od slučajne, ali mora se činiti slučajnom svima koji je presreću. Međutim, ne samo da se mora činiti slučajnom, već ne sme biti ni sumnjiva, kako oni koji se bave otkrivanjem činjenice da

je neki tekst steganografski, ne bi posumnjali. Nasumični setovi znakova od kojih svi spadaju u jedan skup znakova, ali nemaju očigledno značenje, zaista mogu podstaći sumnju.

Ove metode automatski generišu tekstualnu poruku nosioca, pa zato nije potrebna postojeća poruka nosilac. Generirana poruka nosilac koristi tajnu poruku u procesu generiranja. Ovakav algoritam koristi jezičku strukturu i svojstva kao što su: kako se kreiraju rečenice, kakav je format glagola u prošlom ili nekom drugom vremenu itd. Takođe, ove metode koriste gramatiku za proizvodnju odgovarajuće poruke nosioca. U ovoj vrsti tekstualne steganografije dodaje se dodatna složenost (vreme i prostor) kako bi se stvorio ceo paragraf. Ovo zahteva mnogo vremena da se tajna poruka ugradi i izvuče iz poruke nosioca.

Lingvističke metode

Ova metoda se koristi za skrivanje poruke u drugoj poruci u zavisnosti od jezičke strukture poruke nosioca (interpunkcijske oznake) ili semantike reči kao mesta za skrivanje poruke. Dakle, lingvističke metode se mogu podeliti u dve grupe:

- Sintaksna metoda
  Da bi se sakrila poruka, poruka nosilac mora imati interpunkcijske znakove (zarez, tačka itd.) kako bi se sakrila tajna poruka iza njih. Ovi interpunkcijski znakovi su identifikacioni znakovi za tajnu poruku. Ovakva tehnika ima nedostatke, jer veličina podataka zavisi od broja interpunkcijskih znakova poruke nosioca.
- Semantička metoda
  Ova vrsta steganografije koristi sinonime svake reči da bi sakrila poruku. Vrši se pretraživanje sinonima za svaku reč u tajnoj poruci za generisanje izlazne poruke. Ova tehnika se smatra neuspešnom u zaštiti poslate poruke kada neko sa strane pokušava pronaći originalnu poruku zamenivši svaku reč izvornom koristeći semantičke algoritme.

Da bi se rešio problem detekcije neleksičkih sekvenci, stvarne stavke iz rečnika mogu se koristiti za kodiranje jednog ili više bitova informacija po reči. Ovo može uključivati knjigu kodova koja sadrži mapiranja između leksičkih stavki i nizova bitova, ili bi same reči (dužina, slova itd.) mogle da kodiraju skrivene informacije. Međutim, kod ova dva rešenja postoje problemi. Niz reči bez semantičke strukture i bez vidljivog semantičkog odnosa mogu da otkriju i ljudi i računari.

Zbog pomenutih nedostataka je predloženo da se bezkontekstne gramatike (CFG-ovi) koriste kao osnova za stvaranje steganografskih tekstova. Pošto je tekst generisan direktno iz gramatike, osim ako gramatika nije sintaktički pogrešna, zagarantovano je da je tekst sintaktički tačan. Pored toga, stablo stvoreno pomoću CFG-a prirodno se koristi za kodiranje bitova. Strukture stabla se koriste kao optimizacija struktura podataka u mnogim oblastima računarske nauke, od programskih prevodilaca do algoritama za sortiranje. U najjednostavnijoj šemi, u bilo kojem trenutku gde postoji grana u sintaksnom stablu, leva grana može označavati nulu, a desna grana može označavati jedinicu.

Postojeći pristupi

Postoje različite studije koje se odnose na skrivanje podataka u tekstu. U nastavku su prikazane neke od njih. Ono što je zajedničko ovim pristupima steganografije teksta je to da se odnose ili na promenu formata teksta ili na promenu značenja.

Metoda akronima

Ova metoda se odnosi na zamenu reči svojim akronimom. Na ovaj način se može sakriti mala količina podataka. Knjiga kodova sadrži reči i njihove odgovarajuće akronime (tabela 1). Da bi se sakrio bit 0, koristi se cela reč, a za bit 1 se koristi njen akronim.

Akronim	Značenje
218	Too late
ASAP	As soon as possible
C	See
CM	Call me
F2F	Face to face

Tabela 1 – Tabela akronima

Pored same metode akronima, postoje različite metode koje koriste sličan princip. Poruka nosilac se npr. može transformisati tako da sadrži namerne gramatičke ili slovne greške. Takođe mogu se koristiti i skraćenice za određene reči ili nazive.

Metoda promene spelovanja

Ova metoda koristi osobinu britanskog i američkog engleskog jezika, kod kojih se neke reči drugačije speluju. Kodiranje se obavlja na sličan način kao u prethodnoj metodi.

Američko spelovanje	Britansko spelovanje
Favorite	Favourite
Criticize	Criticise
Fulfill	Fulfil
Center	Centre

Tabela 2 – Tabela spelovanja

Semantička metoda

Kod ove metode se koriste sinonimi reči za skrivanje informacija. Međutim, ovaj način zamene određenih reči može dovesti do promene značenja teksta. Pored toga, mala količina podataka se može sakriti ovom metodom.

Reč	Sinonim
Big	Large
Small	Little
Smart	Intelligent

Tabela 3 – Tabela sinonima

Metoda blanko znakova

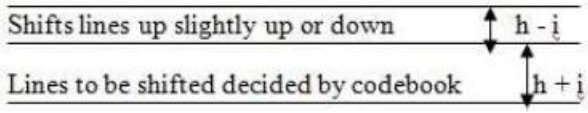
Kod ove metode se koristi princip dodavanja jednog ili dva blanko znaka nakon svake rečenice u tekstu za kodiranje nule, odnosno jedinice. Međutim, na ovaj način se može sakriti mala količina podataka. Varijacija ove metode bi bila ubacivanje blanko znakova na kraju svake linije teksta. Na primer mogu se koristiti dva znaka za kodiranje jednog bita po liniji ili četiri za kodiranje dva bita po liniji itd. Treća opcija je dodavanje blanko znakova nakon svake reči. Na sličan način kao kod prethodnih slučajeva, jedan blanko znak kodira jedinicu, a dva kodiraju nulu.

Sintaksna metoda

Ova metoda za skrivanje podataka koristi sintaksu ili format teksta. Interpunkcijski znakovi kao što su na primer tačka (.) i zarez (,) itd., se na određenim mestima ubacuju u tekst za ovu svrhu. Na ovaj način se kodiraju bitovi 0 i 1. Međutim, problem ove metode je što zahteva identifikaciju tačnih mesta za umetanje interpunkcijskih znakova. Stoga, treba biti oprezan u korišćenju ove metode, jer čitaoci mogu primetiti nepravilnu upotrebu interpunkcijskih znakova.

Metoda pomeranja linija

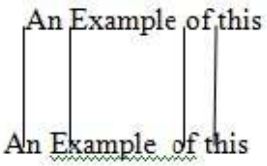
Kod ove metode linije teksta su vertikalno pomerene (na primer 1/400 inča na gore ili dole), što je posebno korisno za štampane tekstove. Označena linija ima dve neoznačene kontrolne linije, po jednu sa obe strane za detekciju smera kretanja obeležene linije. Da bi se sakrio bit 0, linija se pomera prema gore, a da bi se sakrio bit 1, linija se pomera prema dole. Utvrđivanje da li je linija pomerena na gore ili na dole vrši se merenjem udaljenosti centroida obeležene linije i njenih kontrolnih linija. Ako se tekst ponovo upiše ili ako se koristi program za prepoznavanje znakova (Optical Character Recognition - OCR), skrivene informacije će biti uništitene. Takođe, udaljenosti se mogu posmatrati korišćenjem posebnih instrumenata za procenu rastojanja.



Slika 3 – Metoda pomeranja linija

Metoda pomeranja reči

Ovom metodom se manipuliše horizontalnim rastojanjem reči pomoću razmaka. Takođe treba istaći da ova metoda zahteva više vremena, a rezultati ostaju vidljivi ljudskom oku. Tajna poruka se skriva pomeranjem reči horizontalno, tj. levo ili desno da bi se kodirao bit 0 i 1, respektivno. Pomeranje reči otkriva se korelacijskom metodom koja profil tretira kao talasni oblik i odlučuje da li je nastao iz talasnog oblika čiji je srednji blok pomeren ulevo ili udesno. Ovu metodu je teže prepoznati, jer je promena rastojanja između reči zbog poravnanja prilično uobičajena. Međutim, ako neko zna algoritam udaljenosti, on može uporediti stego tekst sa algoritmom i dobiti skriveni sadržaj koristeći razliku. Takođe, prepisivanjem teksta ili korišćenjem OCR programa uništavaju se skrivene informacije.



Slika 4 – Metoda pomeranja reči

Metoda kodiranja karakteristika

Kod ove metode, sintaksa je modifikovana da bi stvorila redundantnosti. Primer za to je širenje ili skraćivanje slova u odnosu na njihove dimenzije. Atributi poput boja se takođe koriste za prikrivanje podataka. U ovoj metodi, poruka je skrivena promenom jedne ili više karakteristika teksta. Analizator pregledava dokument i odabira sve funkcije koje može da koristi za sakrivanje podataka. Na primer, tačka u slovima *i* i *j* može biti pomerena, dužina crtice slova *f* i *t* može se menjati ili se može vršiti produženje ili skraćivanje visine slova *b, d, h,* itd. Važno je obratiti pažnju na strukturu slova jezika kako bi se odabrao odgovarajući način kodiranja. U engleskom jeziku samo slova *i* i *j* imaju tačke, pa bi količina informacija koje se mogu zapamtiti bila jako mala. S druge strane ovaj način bi bio pogodan za arapski jezik koji ima 13 slova sa tačkama od ukupno 26 ili persijsi jezik koji ima 22 slova sa tačkama od ukupno 32. Mana ove metode je u tome što se, ako se koristi OCR program ili ako se vrši ponovno kucanje, skriveni sadržaj se uništava.

Metoda zasnovana na krivama

Ova metoda manipuliše oblikom slova, svrstavajući ih u dve grupe na osnovu njihove strukture. Prvu grupu čine slova koja sadrže krive linije, a drugu ona koja ne sadrže takve linije.

Grupa	Opis	Bit	Slova
A	Sa krivama	0	B,C,D,G,J,O,P,Q,R,S,U
B	Bez krivih	1	A,E,F,H,I,K,L,M,N,T,V,W,X,Y,Z

Tabela 4 – Kodiranje slova zasnovano na krivama

Metoda zasnovana na postojanju vertikalne linije

Ova metoda je slična prethodnoj. Slova se dele u dve grupe na osnovu toga da li njihov oblik sadrži vertikalnu parvu liniju.

Grupa	Opis	Bit	
A	Sa vertikalnom linijom	0	B,D,E,F,H,I,J,K,L,M,N,P,R,T,Y
B	Bez vertikalne linije	1	A,C,G,O,Q,S,U,V,W,X,Z

Tabela 5 – Kodiranje slova zasnovano na postojanju vertikalne linije

Metoda četvorostruke kategorizacije

Ova metoda deli slova u četiri grupe na osnovu toga da li imaju krive linije, horizontalne prave linije, jednu vertikalnu pravu liniju ili dve dijagonalne linije.

Grupa	Opis	Bit	
A	Sa krivama	00	C,D,G,O,Q,S,U
B	Sa poprečnom crtom	01	A,B,E,F,H,P,R
C	Sa jednom vertikalnom linijom	10	I,J,K,L,T,Y
D	Sa dijagonalnim linijama	11	M,N,V,W,X,Z

Tabela 6 – Kodiranje slova zasnovano na četvorostrukoj kategorizaciji

HTML metoda

HTML i XML datoteke se takođe mogu koristiti za skrivanje bitova. Ako postoje različiti početni i završni tagovi, to se tumači kao bit 0, ako se koristi jedan tag za početak i kraj, tada se tumači bit 1. U drugoj tehnici, bit 0 predstavljen je nedostatkom praznog prostora u oznaci, a bit 1 predstavljen je postavljanjem razmaka unutar oznake. U nastavku je dat primer kodiranja.

Stego ključ:
<img></img> -> 0
<img/> -> 1

Stego podaci:
<img src=g1.jpg></img>

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

Na osnovu stego ključa, sakriveni bitovi su: 01110.

CSS (Cascading Stile Sheet)metoda

Ova metoda šifrira poruku koristeći RSA kriptosistem sa javnim ključem, a šifrirani tekst se zatim ugrađuje u Cascading Stile Sheet (CSS) koristeći kraj linije na svim CSS stilovima, neposredno nakon oznake tačka-zarez (;). Prostor nakon tačka-zareza sadrži bit 0, a tabulator nakon tačka-zareza sadrži bit 1.

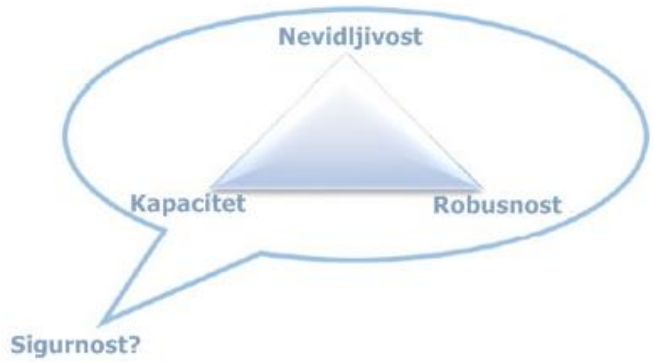
U studiji su analizirani različiti algoritmi skrivanja teksta, a rezultati su dati u tabeli 7. Mogu se uočiti velike razlike u kapacitetu, a pored toga i vreme ugradnje tajne poruke kod različitih algoritama se razlikuje.

Steganografska metoda	Veličina poruke (bajt)	Veličina nosioca (bajt)	Broj sakrivenih karaktera (bajt)	Vreme potrebno za sakrivanje (ms)
Postojanje krivih	800	2640	172	37,996
Postojanje vertikalne linije	800	2640	161	27,533
Četvorostruka kategorizacija	800	2640	145	26,562
Blanko znakovi	800	2640	58	20,825
Kodiranje karakteristika	800	2640	66	18,180

Tabela 7 – Upoređivanje algoritama

Kriterijumi skrivanja teksta

Mnogo je stvari koje treba uzeti u obzir prilikom dizajniranja algoritama za skrivanje teksta. Međutim, osnovni kriterijumi predstavljaju nevidljivost, kapacitet ugradnje, robusnost i sigurnost. Takođe, odabir poruke nosioca je važna karakteristika. U zavisnosti od mrežnih aplikacija, potrebno je tražiti kompromis za zadovoljenje kriterijuma u bilo kojoj tački trougla kriterijuma kao što je prikazano na slici 5 .



Slika 5 – Evaluacioni kriterijumi algoritama skrivanja teksta

Nevidljivost

Kvantifikovanje napadača ili njegove sposobnosti da otkrije poruku ili njeno postojanje naziva se nevidljivost (neprimetnost / detektabilnost / transparentnost) . To znači da trag ugrađivanja tajne poruke u poruku nosioca mora biti nevidljiv i treba da ukloni mogućnost

sumnje koje se javlja na osnovu kod ljudskog vida. Drugim rečima, nevidljivost se odnosi na to koliko je perceptivnih modifikacija izvršeno u poruci nosiocu nakon ugradnje tajne poruke. To se praktično ne može meriti brojčano. Najbolji način analize stepena nevidljivosti je poređenje varijacija poruke nosioca i stego poruke, tj. sa i bez tajne poruke.

Kapacitet ugradnje

Broj tajnih bitova koji se mogu ugraditi u poruku nosioca (CM) naziva se kapacitetom ugradnje ili korisnim opterećenjem. Ova karakteristika se može numerički izmeriti u jedinicama bita po lokaciji (BPL) ili karaktera po lokaciji (CPL). Lokacija znači promenljivu karakteristiku (znak / reč) koja se može smatrati ugradljivom lokacijom (EL) u poruci nosiocu, kao što je prostor između reči, prostor posle posebnih znakova itd. Iako algoritam tekstualne steganografije može da pruži veći kapacitet ugradnje, neće biti efikasan ako duboko izmeni poruku nosioca. Kapacitet se računa po formuli:

$EC_{CM} = BPL\ x\ EL_{CM}$ ili $EC_{CM} = CPL\ x\ EL_{CM}$

Robusnost distorzije

Nad stego porukom se može dogoditi više napada dok se ona prenosi kanalima gde može biti izložena opasnosti koja može uništiti tajnu poruku. Štaviše, napadači mogu pokušati manipulisati tajnom porukom, a ne ukloniti je. Stoga se bilo koja vrsta izobličenja može pojaviti namerno ili čak nenamerno u stego poruci. Robustan algoritam skrivanja teksta čini tajnu poruku izuzetno teškom za promenu ili uništavanje.

Sigurnost

Postoji određeni nivo sigurnosti koji sprečava napadače da vizuelno otkriju tajnu poruku ili da je uklone iz stego poruke. Ova mera zavisi od tri druga kriterijuma: nevidljivosti, kapaciteta ugradnje i robusnosti distorzije. Efikasan steganografski algoritam mora da pruža optimalan kompromis među ovim kriterijumima. U savremenim tehnikama skrivanja teksta, kriptosistem se može koristiti za zaštitu tajnih bitova od napada dekodiranja. U praksi se koristi funkcija enkripcije da bi se osigurali bitovi tajne poruke pre nego što se ugrade u poruku nosioca i da bi se izmeno redosled tajnih bitova tako da ih može izvući samo odgovarajuća funkcija dešifriranja. Verovatnoća dekodiranja (DP) je verovatnoća dekodiranja originalnih bitova pomoću napada nagađanjem. Pretpostavimo da napadač nagađa da poruka može sadržati tajnu poruku (npr. on nema pojma o pristupu koji je korišćen za prikrivanje). Štaviše, napadač može pokušati da dekodira stego poruku koristeći konvencionalne pristupe ili pogodi bitove (koristeći analizu raspodele verovatnoće) iz nevidljivih simbola ili karakteristika. Pošto se funkcija enkripcije koristi za osiguranje bitova na osnovu tajnog ključa (K), nemoguće je dekodirati originalni bit iz šifriranog bita bez tajnog ključa i odgovarajuće funkcije dešifriranja.

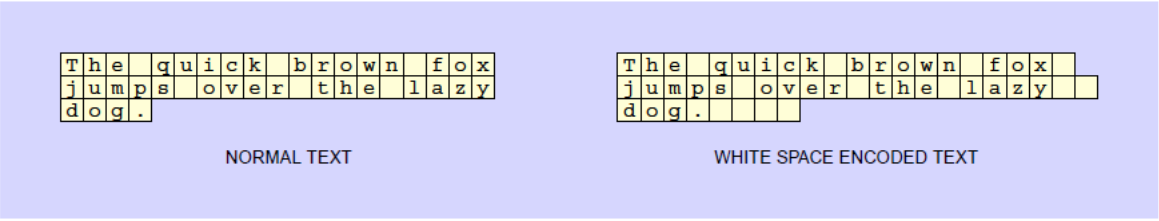
Metode blanko znakova (Open space method)

Sakrivanje informacija unutar praznih prostora ima potencijal, jer ljudi teško mogu znati za postojanje skrivenih bitova. Postoje dva razloga zbog kojih manipulacija belim prostorom (razmakom) daje korisne rezultate. Prvo, promena broja praznih mesta ima malo šanse da promeni značenje fraze ili rečenice. Drugo, čitalac verovatno neće primetiti male izmene belog prostora. Postoji više načina korišćenja belog prostora za kodiranje podataka. Različite metode koriste razmake između rečenica, razmake na kraju reda i razmake između reči u odabranom tekstu.

Jedna od najjednostavnijih metoda kodira binarnu poruku u tekst postavljanjem jednog ili dva razmaka nakon svakog završnog znaka, kao što je na primer tačka ili zarez koji se često javljaju u proznim tekstovima ili tačka-zarez koji se često javlja u C, C++, Java… kodovima. Jedan razmak kodira bit 0, dok dva razmaka kodiraju bit 1 ili obrnuto.

Ova metoda ima niz urođenih problema. Neefikasna je i zahteva mnogo teksta da bi kodirao vrlo malo bitova. Jedan bit po rečenici jednak je brzini prenosa podataka od približno jednog bita po 160 bajtova, pod pretpostavkom da su rečenice u proseku dve linije teksta od po 80 znakova. Sposobnost kodiranja zavisi od strukture teksta. Nekim tekstovima, kao što je poezija slobodnih stihova, nedostaju dosledni ili dobro definisani terminalni znakovi.

Druga metoda iskorišćavanja belog prostora za kodiranje podataka bi bila umetanje razmaka na kraj linija. Podaci se kodiraju pomoću unapred određenog broja razmaka na kraju svake linije. Dva razmaka kodiraju jedan bit po liniji, četiri kodiraju dva, osam kodiraju tri, itd. Na ovaj način se drastično povećava količina informacija koja se može kodirati u odnosu na prethodnu metodu. Na slici 6, odabran je odgovarajući tekst, a zatim su dodati razmaci na kraju redova za kodiranje više podataka. Dodatne prednosti ove metode su u tome što se može raditi sa bilo kojim tekstom, a kod čitaoca će proći nezapaženo, jer je ovaj dodatni beli prostor periferni prema tekstu. Kod ove, kao i kod prethodne metode, neki programi, kao što je npr. *Sendmail*, mogu nenamerno ukloniti dodatne razmake. Još jedan problem, koji je jedinstven za ovu metodu je taj što se skriveni podaci ne mogu preuzeti sa štampane kopije.



Slika 6 – Korišćenje blanko znakova na kraju redova

Metoda blanko znakova kod poravnatog teksta [6]

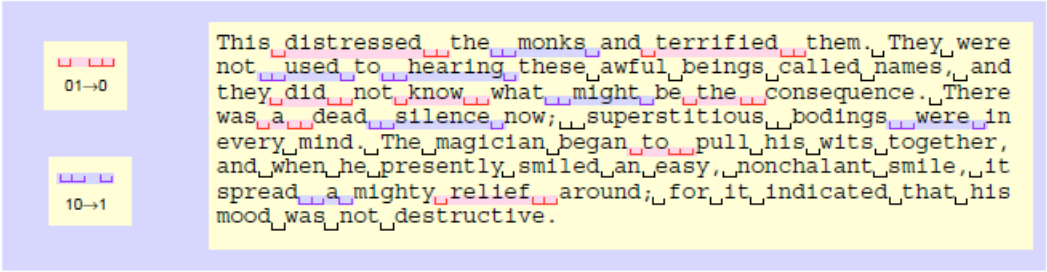
Kod ove metode, podaci se kodiraju kontrolisanjem postavljanja praznih prostora (razmaka) u poravnatom tekstu. Na ovaj način se značajno dobija na kapacitetu u odnosu na dve

prethodno opisane metode. Jedan razmak između reči tumači se kao bit 0, a dva prostora se tumače kao bit 1. Ova metoda rezultuje s nekoliko bitova kodiranih u svakoj liniji. Ideja ove metode je da se originalni i stego tekst ne razlikuju po rasporedu reči po linijama. Zbog ograničenja kod poravnanja teksta, ne može se svaki prostor između reči koristiti kao podatak. Da bi se odredilo koji prostor između reči predstavljaju skrivene bitove podataka, a koji su deo originalnog teksta, može se koristiti Mančester metoda kodiranja. Mančester kodiranje grupiše bitove u skupove od po dva bita, tumačeći 01 kao 1, a 10 kao 0 (slika 7). Nizovi bita 00 i 11 su *null*. Na primer, kodirana poruka 1000101101 je smanjena na poruku 001, dok je 110011 *null* string.

Kapacitet zavisi od poravnatog teksta. U najboljem slučaju bi se mogao iskoristiti svaki razmak u svakoj liniji, pa bi broj sakrivenih bitova u svakoj liniji iznosio:

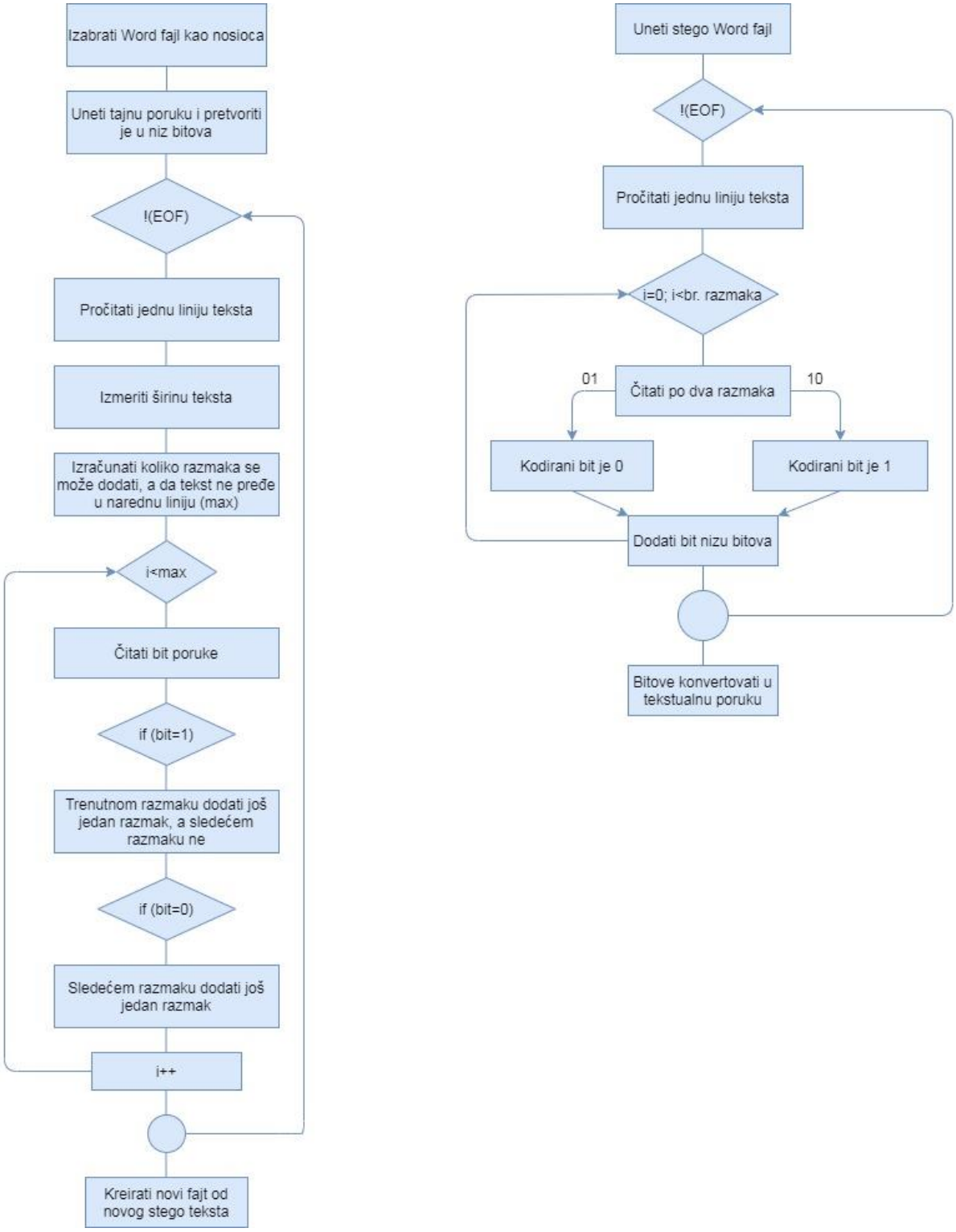
x = *br. reči u liniji – 1* za neparan broj reči, odnosno x = *br. reči u liniji – 2* za paran broj reči.

Kapacitet bi onda iznosio $k_s = \sum_{n=1}^{br.linija} x_n$, gde je x_n broj sakrivenih bitova u jednoj liniji teksta. Kako se koristi Mančester kodiranje, ukupni kapacitet tajne poruke je zapravo još manji. Svaka dva bita kodirane poruke se prevode u jedan bit prave poruke, pa je kapacitet $k = k_s/2$.



Slika 7 – Metoda blanko znakova sa Mančester kodiranjem

Na slici 8 je dat algoritam skrivanja poruke kod ove metode (levo) i algoritam ekstraktovanja poruke (desno). Iako metoda pruža poboljšanje u odnosu na prethodno opisane, i dalje nije svaki razmak iskorišćen. Pored toga se zahteva dodatni korak merenja dužine teksta u svakoj liniji kako bi se utvrdilo koliko razmaka može biti dodato, a da se ne desi prelazak poslednje reči u novu liniju.



Slika 8 – Algoritmi skrivanja i ekstraktovanja poruke kod metode poravnatog teksta

Na slici 9 je prikazan originalni tekst kod koga je korišćeno poravnanje, a na slici 10 rezultat umetanja tajne poruke „ps”.

Gazele su poznate kao brze životinje. Neke mogu da potrče u namasima brzinama od po 100 km/h (60 mph) ili da duže trče brzinom od 50 km/h (30 mph). Gazele uglavnom obitavaju u pustinjama, pašnjacima i savanama Afrike; ali se takođe nalaze u jugozapadnoj i centralnoj Aziji i Indijskom potkontinentu. One imaju tendenciju da žive u stadima i jedu manje grube, lako svarljive biljke i lišće. Gazele su relativno male antilope, pri čemu većina ima stojeću visinu od 60-1.070 cm (2-35 ft) u ramenima, i generalno su žućkaste boje.

Slika 9 – Tekst nosilac

Gazele su poznate kao brze životinje. Neke mogu da potrče u namasima brzinama od po 100 km/h (60 mph) ili da duže trče brzinom od 50 km/h (30 mph). Gazele uglavnom obitavaju u pustinjama, pašnjacima i savanama Afrike; ali se takođe nalaze u jugozapadnoj i centralnoj Aziji i Indijskom potkontinentu. One imaju tendenciju da žive u stadima i jedu manje grube, lako svarljive biljke i lišće. Gazele su relativno male antilope, pri čemu većina ima stojeću visinu od 60-1.070 cm (2-35 ft) u ramenima, i generalno su žućkaste boje.

Slika 10 – Stego tekst

Metode praznog prostora su korisne sve dok je tekst u ASCII (American Standard Character Interchange) formatu, odnosno ne može se primeniti na jezike kod kojih ne postoje prazni prostori. Kao što je gore spomenuto, neki podaci se mogu izgubiti prilikom štampanja teksta. Sakrivanje podataka u papirnoj kopiji postiže se malim odstupanjima u razmaku reči i slova, promenama u početnom položaju slova ili interpunkcije, promenama u obrascima slova itd.

Metoda blanko znakova kod neporavnatog teksta

Ova šema ugradnje primenjena je u prostoru koji se pojavljuje između reči. Glavna prednost ove metode u odnosu na prethodnu je ta što se svaki prazni prostor može iskoristiti za ugradnju bita tajne poruke. Na primer, znak je ekvivalentan prostoru od 8 bitova, pa je za kodiranje jednog znaka potrebno 8 međuprostora, odnosno razmaka.

U algoritmu je potrebno izmeriti i broj praznih prostora u datoteci kako bi se utvrdilo da li izabrani fajl može da se koristiti. Znajući broj znakova i broj razmaka ovo se može lako utvrditi.

Algoritam skrivanja se sastoji iz sledećih koraka:

- Prvo se unese tajna poruka koja treba biti sakrivena.
- Zatim se izabere datoteka koja sadrži tekst (eng. cover text) u kojem treba biti sakrivena tajna poruka.

- Generiše se tekstualna datoteka (stego datoteka) u kojoj je skrivena tajna poruka.

Ekstraktovanje skrivene poruke je jednostavno i obavlja se detektovanjem jednog razmaka koji reprezentuje nulu ili dva spojena razmaka koji reprezentuju jedinicu.

Kapacitet je direktno proporcionalan broju reči u tekstu nosiocu. Dakle, iznosi:

$k = br.\ reči - 1$

Na slici 11 je prikazan originalni tekst, a na slikama 12 i 13 tekst sa umetnutom tajnom porukom. Kod ove metode se može raditi sa tekstom sa ili bez poravnanja. U oba slučaja, zbog dodavanja dodatnih razmaka, stego tekst ne mora nužno biti isti kao i originalni u pogledu rasporeda reči po linijama.

Gazele su poznate kao brze životinje. Neke mogu da potrče u namasima brzinama od po 100 km/h (60 mph) ili da duže trče brzinom od 50 km/h (30 mph). Gazele uglavnom obitavaju u pustinjama, pašnjacima i savanama Afrike; ali se takođe nalaze u jugozapadnoj i centralnoj Aziji i Indijskom potkontinentu. One imaju tendenciju da žive u stadima i jedu manje grube, lako svarljive biljke i lišće. Gazele su relativno male antilope, pri čemu većina ima stojeću visinu od 60—1.070 cm (2—35 ft) u ramenima, i generalno su žućkaste boje.

Slika 11 – Tekst nosilac (neporavnat)

Gazele su poznate kao brze životinje. Neke mogu da potrče u namasima brzinama od po 100 km/h (60 mph) ili da duže trče brzinom od 50 km/h (30 mph). Gazele uglavnom obitavaju u pustinjama, pašnjacima i savanama Afrike; ali se takođe nalaze u jugozapadnoj i centralnoj Aziji i Indijskom potkontinentu. One imaju tendenciju da žive u stadima i jedu manje grube, lako svarljive biljke i lišće. Gazele su relativno male antilope, pri čemu većina ima stojeću visinu od 60—1.070 cm (2—35 ft) u ramenima, i generalno su žućkaste boje.

Slika 12 – Stego tekst (neporavnat)

Gazele su poznate kao brze životinje. Neke mogu da potrče u namasima brzinama od po 100 km/h (60 mph) ili da duže trče brzinom od 50 km/h (30 mph). Gazele uglavnom obitavaju u pustinjama, pašnjacima i savanama Afrike; ali se takođe nalaze u jugozapadnoj i centralnoj Aziji i Indijskom potkontinentu. One imaju tendenciju da žive u stadima i jedu manje grube, lako svarljive biljke i lišće. Gazele su relativno male antilope, pri čemu većina ima stojeću visinu od 60-1.070 cm (2-35 ft) u ramenima, i generalno su žućkaste boje.

Slika 13 – Stego tekst (poravnat)

Izazovi vezani za skrivanje podataka u tekstualnim datotekama na ovaj način uključuju:

- Nizak kapacitet skladištenja teksta u tekstualnoj datoteci.
- Umetanje dodatnih prostora za predstavljanje informacija rezultuje povećanjem veličine stego datoteke.
- Promena jednog bita u jednom bajtu rezultuje u potpuno različitom ACII kodu koji može / ne mora imati nikakvu relevantnost sa tekstualnim sadržajem.

Metode karaktera koji nemaju tekstualni trag (Zero-width method) [8,9]

Tehnike zasnovane na belom (praznom) prostoru mogu da koriste posebne *Unicode* karaktere za ugrađivanje tajnih bitova u poruku nosioca, na primer: između reči, na krajevima rečenica i tako dalje. U praksi, ove tehnike pružaju visok nivo nevidljivosti, nizak kapacitet ugrađivanja i skromnu robusnost protiv vizuelnih napada. Štaviše, mogu se primeniti u tekstovima na različitim jezicima.

Tehnike zasnovane na *zero-widht* znacima (znaci nulte širine - ZWC) koriste ZWC *Unicode* znakove za umetanje bitova tajne poruke u poruku nosioca. ZWC-ovi se inače koriste za pružanje određenih entiteta kao što je Zero-width Joiner (ZWJ), koji spaja dva podržana znaka u određenim jezicima. Sa stanovišta obrade teksta, ZWC karakteri nemaju tekstualni trag (napisan simbol), niti širinu i mogu se ugrađivati na različite lokacije kroz tekst. Ovi pristupi se mogu koristiti u višejezičnim tekstovima i na različitim platformama za obradu teksta kao što su društvene mreže, e-pošta, SMS, itd. Pošto ZWC-ovi imaju nevidljive tragove u tekstu, oni se mogu ugraditi koristeći maksimalan broj slova u kanalu za komunikaciju (npr. SMS, Facebook, itd.). U praksi, ovakvi pristupi pružaju visoku nevidljivost, visok kapacitet ugrađivanja i veću robusnost protiv strukturalnih napada.

Na nekim platformama društvenih mreža, ako se koristi standard *Unicode* za obradu digitalnih tekstova na različitim jezicima, ZWC-ovi predstavljaju nevidljive pisane simbole. U suprotnom, oni mogu prikazati neke neobične simbole. Na osnovu eksperimenata, utvrđeno je da je Gmail blokirao znak “U + 200B”, a Apple iOS ne dozvoljava prenos “U + 200D”.

ZWC i blanko karakter metoda

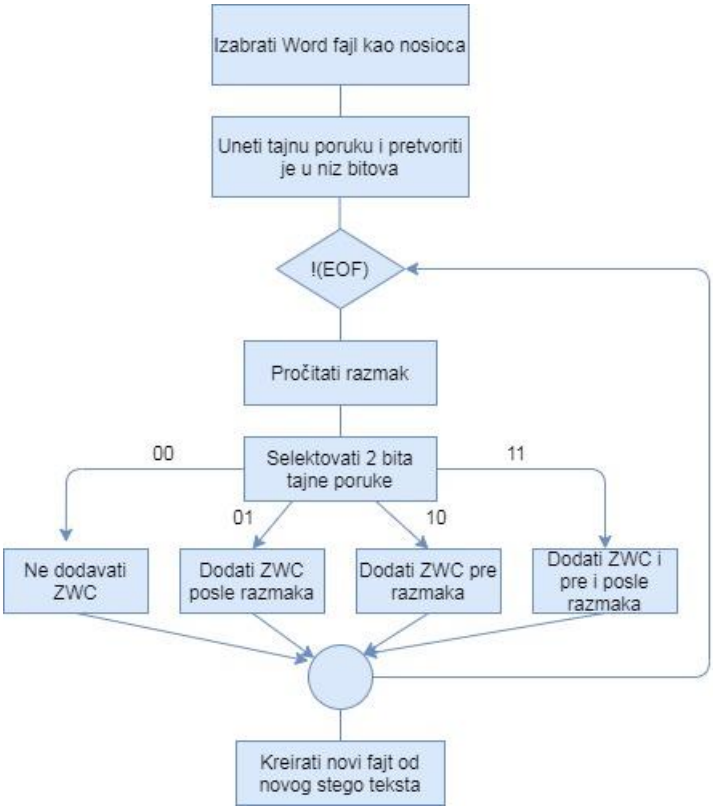
U ovom algoritmu [8] se koristiti Zero-Width Character. ZWC je Unicode znak (U + 200B), koji ne zauzima nikakav prostor ili formatiranje datoteka. Dodavanjem ZWC-a pre i posle praznog prostora mogu se sakriti podaci. Microsoft Word može prebrojati broj znakova u bilo kojoj datoteci bez i sa razmacima, a nakon dodavanja ZWC-a neće se povećati broj karaktera ni u kom slučaju.

Kod ove metode kapacitet iznosi:

$k = br.\ razmaka*2$, jer svaki razmak može sakriti 2 bita.

Najpogodnija datoteka za sakrivanje je ona koja ima što veći *space ratio*, gde on iznosi: *space ratio = br. razmaka / br. karaktera.*

Algoritam za ovu metodu je prikazan na slici 14.



Slika 14 – Algoritam skrivanja poruke kod ZWC metode

Ekstraktovanje poruke se obavlja traženjem razmaka i detektovanja postojanja ZWC karaktera. Na osnovu toga gde se ovaj karakter nalazi dekodiraju se bitovi 00, 01, 10 i 11.

Na slici 15 je dat izgled stego teksta nakon primene algoritma, gde je sakrivena poruka „pass“. Za tekst nosioca je korišćen tekst sa slike 9. U ovom slučaju se golim okom nikako ne može uočiti razlika ova dva teksta.

Gazele su poznate kao brze životinje. Neke mogu da potrče u namasima brzinama od po 100 km/h (60 mph) ili da duže trče brzinom od 50 km/h (30 mph). Gazele uglavnom obitavaju u pustinjama, pašnjacima i savanama Afrike; ali se takođe nalaze u jugozapadnoj i centralnoj Aziji i Indijskom potkontinentu. One imaju tendenciju da žive u stadima i jedu manje grube, lako svarljive biljke i lišće. Gazele su relativno male antilope, pri čemu većina ima stojeću visinu od 60-1.070 cm (2-35 ft) u ramenima, i generalno su žućkaste boje.

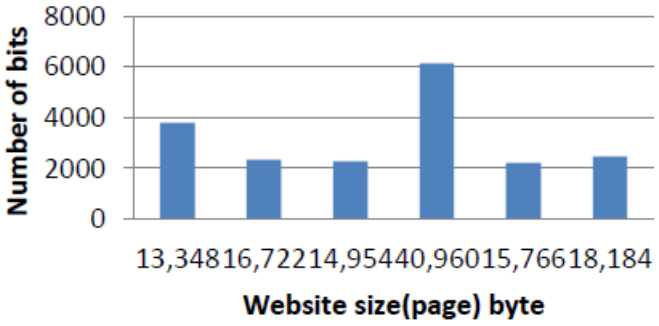
Slika 15 – Stego tekst ZWC metode

ZWC algoritam ima sledeće prednosti:

- Formatirana datoteka neće biti promenjena.
- Može se primeniti za bilo koji kod (Unicode, ASCII). Drugim rečima, ovaj algoritam predstavlja opšti oblik za bilo koji jezik.
- Veličina datoteke neće biti znatno promenjena.

Problem ovih pristupa, koji koriste blanko i posebne nevidljive karaktere, je nedostatak robusnosti zbog nepostojanja dodatnih slojeva sigurnosti. Ako neko oseti da je primenjana ovakva metoda, bio bi u stanju da izvuče skrivene informacije.

U ovoj studiji je izvršeno skrivanje poruka u tekstovima različitih veb stranica. Na slici 16 je dat rezultat koji prikazuje odnos veličine stranice i broja sakrivenih bitova.



Slika 16 – Odnos veličina veb stranica i broja sakrivenih bitova

Metoda korišćenja nevidljivih simbola

Algoritam predstavljen u ovoj studiji [9] skriva podatke unutar *word* datoteke bez bilo kakvih promena u svojstvima datoteke poput veličine, sadržaja i formata datoteke. Predloženi algoritam koristi nevidljive simbole da sakrije četiri bita između karaktera datoteke nosioca, što poboljšava kapacitet skrivanja u odnosu na prethodne algoritme. Štaviše, nikakve promene u obliku reči ili obliku slova se ne vrše. Pored toga, predloženi algoritam izbegava sumnje i bilo kakvu primetnost za stegoanalizator, što će zauzvrat poboljšati robusnost algoritma. Umetanje nekog od simbola iz tabele varijacija nakon svakog slova omogućava sakrivanje četiri bita. Uglavnom se koriste *Right remark* (200E), *Left remark* (200F), *Zero width joiner* (200D) i *Zero width non-joiner* (200C), koji se ugrađuju u datoteku nosioca. U ovoj tehnici, različite varijacije se mogu koristiti za predstavljanje skrivenih bitova za ukupno 16 različitih kodova, kao što je prikazano u tabeli 8.

Right Remark	Left Remark	ZWJ	ZWNJ	Sakriveni kod
X	X	X	X	0000
X	X	X		0001
X	X		X	0010
X	X			0011
X		X	X	0100
X		X		0101
X			X	0110
X				0111
	X	X	X	1000
	X	X		1001
	X		X	1010
	X			1011
		X	X	1100
		X		1101
			X	1110
				1111

Tabela 8 – Kodiranje poruke znakovima

Slika 17 predstavlja korake sakrivanja podataka kada se koriste tri ulaza - datoteka nosilac, skrivena poruka i stego ključ. Glavna svrha stego ključa je da promeni kodiranje bitova simbolima. Drugim rečima, bit 0 predstavlja odsustvo karaktera, dok se drugo stanje predstavlja bitom 1. U sledećem koraku se kreira tabela simbola u zavisnosti od stego ključa. Zatim se ubacuju četiri bita iz skrivenih podataka nakon svakog slova u datoteku nosioca. Primalac može da preuzme skrivene podatke čitanjem datoteke nosioca i pomoću stego ključa da napravi tabelu simbola. Čitanje simbola nakon svakog slova i njihovo podudaranje sa tabelom simbola omogući će primaocu da izvuče skrivene podatke.

Kapacitet kod ovog algoritma iznosi:

*k = br. karaktera*4*

Slika 17 – Algoritam skrivanje poruke kod metode korišćenja nevidljivih simbola

Na slici 18 je prikazan rezultat skrivanja tajne poruke „pass“ korišćenjem ove metode.

Gazele su poznate kao brze životinje. Neke mogu da potrče u namasima brzinama od po 100 km/h (60 mph) ili da duže trče brzinom od 50 km/h (30 mph). Gazele uglavnom obitavaju u pustinjama, pašnjacima i savanama Afrike; ali se takođe nalaze u jugozapadnoj i centralnoj Aziji i Indijskom potkontinentu. One imaju tendenciju da žive u stadima i jedu manje grube, lako svarljive biljke i lišće. Gazele su relativno male antilope, pri čemu većina ima stojeću visinu od 60-1.070 cm (2-35 ft) u ramenima, i generalno su žućkaste boje.

Slika 18 – Stego tekst

Algoritam ima mnogo prednosti u odnosu na druge algoritme. Na primer, ovaj algoritam se može primeniti na bilo koji jezik, bez obzira da li se radi o *Unicode* ili ASCII kodovima, pri čemu se drugi algoritmi, npr. kao što su oni koji koriste razmake mogu primeniti na samo jezike kod kojih postoje razmaci između reči. Štaviše, nema potrebe za posebnim softverom ili opremom da bi se sakrili podaci i izdvojili. Algoritam ne menja format datoteke, jer upotrebljeni simboli ne utiču na format slova. Shodno tome, ovaj algoritam poboljšava funkciju transparentnosti koja je jedan od ključnih ciljeva stenografije.

Kompresija tajne poruke

Kako je količina podataka koji se mogu sakrati u nekoj datoteci važna osobina steganografije, neke studije predlažu korišćenje algoritama za kompresiju, kako bi broj skrivenih bitova bio veći.

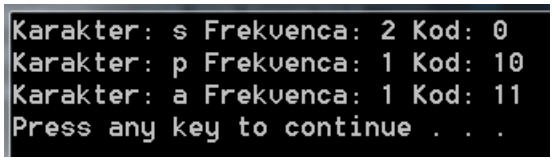
Mogu se koristiti algoritmi kompresije bez gubitaka, kao što su Hafmanovo kodiranje (eng. Huffman coding), LZW (Lempel-Ziv-Welch), aritmetičko kodiranje i tako dalje. Ovi algoritmi su pogodni tokom kodiranja u metodama zasnovanim na formatu za poboljšanje kriterijuma kapaciteta ugradnje tajne poruke. Efikasan algoritam skrivanja teksta trebalo bi da obezbedi optimalan kompromis među tri osnovna kriterijuma za postizanje određenog nivoa.

Hafmanovo kodiranje

Hafmanovo kodiranje je algoritam kompresije podataka bez gubitaka zasnovan na učestalosti pojavljivanja podataka. Ideja je dodeliti kodove promenljive dužine znakovima na osnovu frekvencija pojavljivanja odgovarajućih znakova u tekstu. Najčešći znak dobija najkraći kod, a najređi znak dobija najduži kod. Generisani kodovi su prefiksni kodovi, tj. kod dodeljen jednom znaku nije prefiks koda dodeljenog bilo kom drugom znaku. Hafmanovo stablo je izgrađeno od ulaznih znakova, a zatim se kodovi dodeljuju znakovima prolazeći kroz stablo.

Algoritam se zasniva na kreiranju stabla na osnovu frekvencija pojavljivanja karaktera u tekstu. Listovi stabla su karakteri i kod svakog od njih se dobija čitanjem bitova od korena stabla do lista, pri čemu leva grana ima vrednost 0, a desna vrednost 1.

Prilikom sakrivanja poruke „pass“ korišćenjem Hafmanovog kodiranja, ona se sastoji od 6 bitova: 101100. Na slici 19 je prikazano koji kod je dodeljen kom slovu. Na ovaj način se može uštedeti dosta prostora, naročito kod tekstova koji se sastoje od karaktera koji se ponavljaju veliki broj puta. U ovom konkretnom slučaju tajna poruka bez Hafmanovog kodiranja sadrži 32 bita (po 8 bitova za svako slovo), pa je smanjenje više od 5 puta.

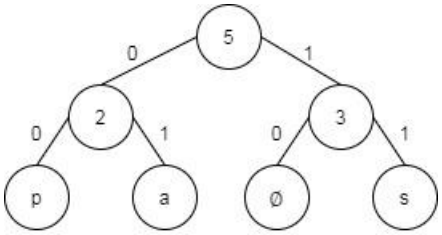


Slika 19 – Kodiranje teksta „pass“

Prilikom ovakve kompresije neophodno je postaviti neki graničnik koji će odrediti koliko bitova će biti dekodirano prilikom ekstrakcije tajne poruke iz stego teksta. Ovo je neophodno, jer kompresijom, karakteri nemaju istu dužinu koda, niti će isti karakter imati isti kod u različitim tekstovima. Takođe, Hafmanov algoritam se ne može primeniti uz prethodno opisane algoritme koji sakrivaju poruku uzimajući grupe njenih bitova (po 2 ili 4 bita). Svakako, kod algoritma koji sakriva po jedan bit u svakom razmaku, Hafmanovo kodiranje se

može primeniti i korisno je, jer omogućava sakrivanje više bitova, odnosno povećava kapacitet algoritma.

Za potrebe stego algoritma koji koristi Hafmanovu kompresiju, uveden je terminalni karakter koji se nadovezuje na tajnu poruku i tako predstavlja njen sastavni deo i određuje njen kraj. Na slici 20 je prikazano Hafmanovo stablo za poruku „pass“ sa terminalnim karakterom, a na slici 21 kodovi karaktera u tom slučaju.



Slika 20 – Hafmanovo stablo

```
Karakter: p Frekvenca: 1 Kod: 00
Karakter: a Frekvenca: 1 Kod: 01
Karakter: null Frekvenca: 1 Kod: 10
Karakter: s Frekvenca: 2 Kod: 11
```

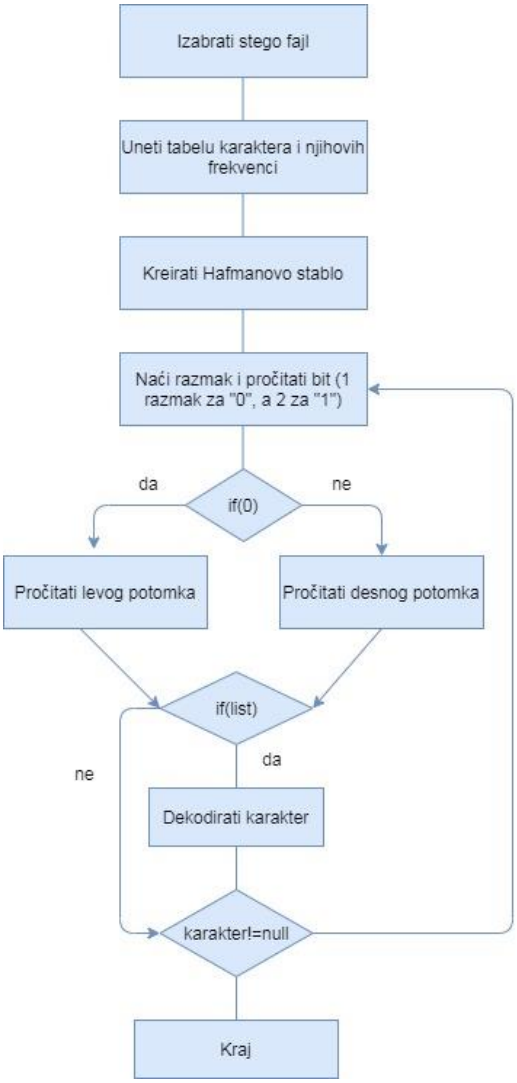
Slika 21 – Kodiranje teksta „pass“ sa terminalnim karakterom

Na slici 22 je prikazan stego algoritam koji kompresuje poruku i zatim je ugrađuje u tekst nosioca uz pomoć algoritma zasnovanog na blanko znakovima.

Slika 22 – Algoritam skrivanja kod metode sa kompresijom

Ekstraktovanje poruke iz stego datoteke je jednostavno i obavlja se čitanjem svagog bita i prolaženjem kroz stablo kako bi se pronašao karakter koji je dodeljen kodu. Čitanje bitova se

obavlja sve dok se ne dekodira terminalni karakter koji označava da je cela poruka ekstraktovana. Međutim, kod ovakvog pristupa mana je to što je i primaocu potrebno da zna tablicu kodova kako bi mogao da dekodira poruku. Algoritam ekstrakcije poruke je prikazan na slici 23, a na slici 24 rezultat skrivanja poruke „pass“.



Slika 23 – Algoritam ekstrakcije poruke kod metode sa kompresijom

Gazele su poznate kao brze životinje. Neke mogu da potrče u namasima brzinama od po 100 km/h (60 mph) ili da duže trče brzinom od 50 km/h (30 mph). Gazele uglavnom obitavaju u pustinjama, pašnjacima i savanama Afrike; ali se takođe nalaze u jugozapadnoj i centralnoj Aziji i Indijskom potkontinentu. One imaju tendenciju da žive u stadima i jedu manje grube, lako svarljive biljke i lišće. Gazele su relativno male antilope, pri čemu većina ima stojeću visinu od 60-1.070 cm (2-35 ft) u ramenima, i generalno su žućkaste boje.

Slika 24 – Stego tekst

Kompresija po grupama

Iako Hafmanovo kodiranje značajno smanjuje dužinu poruke koja se ugrađuje, nedostatak je to što su kodovi karaktera različite dužine. Kako bi se ovo izbeglo, kompresija po grupama omogućuje ravnomerno smanjenje dužine.

Ovakav algoritam [12] pravi blokove binarnog niza tajne poruke koji imaju 4 bita po bloku kako bi se smanjila veličina niza bitova dizajnirajući rečnik kao što je prikazano u tabeli. 9. Rečnik sadrži svih šesnaest mogućih kombinacija od 4 bita, koji se zatim preslikavaju u dva bita, tako što se uzimaju poslednja 2 bita grupe. Na ovaj način svih šesnaest kombinacija je grupisano u četiri mapirana kombinaciona bloka koja se nazivaju G1, G2, G3 i G4. Određivanje grupe kojoj pripada 2-bitno mapiranje se određuje na osnovu prva 2 bita četvoročlane kombinacije. Na taj način ako kombinacija počinje sa 00, mapiranje pripada grupi G1, ako počinje sa 01, pripada grupi G2 itd. Naziv grupese koristi kao ključ prilikom ugrađivanja poruke i koristic se u procesu ekstrakcije kako bi se dobio 4-bitni blok koji odgovara svakoj 2-bitnoj grupi.

Kombinacije	Mapiranje	Grupa
0000	00	G1
0001	01	G1
0010	10	G1
0011	11	G1
0100	00	G2
0101	01	G2
0110	10	G2
0111	11	G2
1000	00	G3
1001	01	G3
1010	10	G3
1011	11	G3
1100	00	G4
1101	01	G4
1110	10	G4
1111	11	G4

Tabela 9 – Rečnik kompresije po grupama

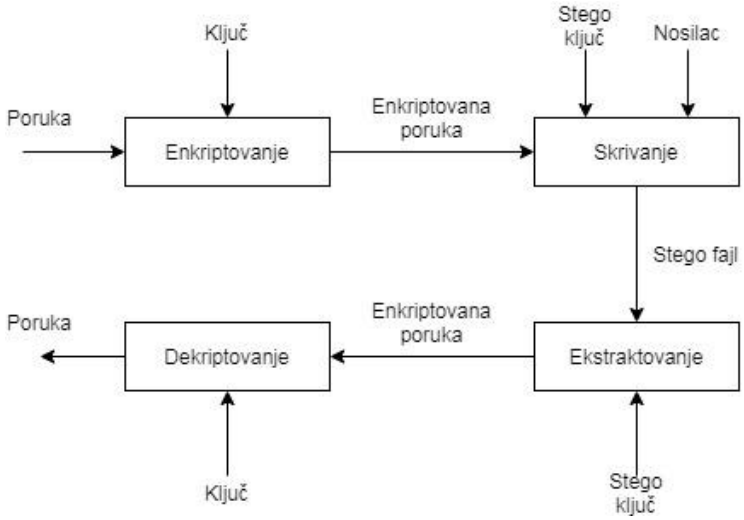
Algoritam, dakle, skenira tajne blokove i smanjuje svaki blok od 4 bita na 2 bita prema ključu i stvara niz sa nazivima grupa. 2 bita koja predstavljaju naziv grupe isključuju se iz procesa ugradnje i samo se desna dva bita koriste da bi se stvorio novi niz, smanjujući veličinu tajnog niza bitova na polovinu.. Stoga je predložena metoda višeslojna arhitektura koja značajno povećava kapacitet, sigurnost i robusnost.

Enkripcija

U steganografiji je sigurnost poruke od najvećeg značaja. Stoga se za poboljšanje sigurnosti tajne poruke može koristiti koncept nekog od šifratora u modelu tekstualne steganografije. Na taj način, čak i ako se otkrije algoritam za umetanje i vađenje podataka, još uvek je potrebno mnogo napora da se poruka dešifruje. Za šifriranje se može koristiti bilo koji simetrični ili asimetrični šifrator (One-Time-Pad, AES, DES…) i pri tom su i pošiljaocu i primaocu potrebni odgovarajući ključevi kako bi se poruka šifrirala i kasnije, ekstraktovana poruka dešifrirala.

Na slici 25 je prikazan predloženi model steganografije teksta. Model se sastoji od četiri bloka:

- Enkriptujuća funkcija, koja enkriptuje poruku pomoću algoritma za šifriranje
- Funkcija skrivanja, koja skriva kodiranu poruku pomoću stego ključa
- Funkcija traženja, koja ekstraktuje sakrivene informacije iz stego fajla, koristeći stego ključ
- Dekriptujuća finkcija, koja dekriptuje ekstraktovanu poruku pomoću tajnog kluča



Slika 25 – Steganografija sa enkripcikom

Što se tiče enkripcije, korišćen je AES algoritam. Nakon toga, šifrirana poruka je kompresovana korišćenjem kompresije po grupama, kako bi se povećao kapacitet ugradnje. Kako se nakon kompresije svaki karakter kodira pomoću 4 bita, primenjen je algoritam nevidljivih simbola koji ugrađuje 4 bita nakon svakog karaktera. Za skrivanje su potrebna 2 stego-ključa i to jedan za kompresiju kako bi se odredili nazivi grupa, a jedan za ugrađivanje poruke kako bi se odredio raspored nevidljivih karaktera (koji karakter kodira koji bit). Rezultat je stego fajl prikazan na slici 26.

Gazele su poznate kao brze životinje. Neke mogu da potrče u namasima brzinama od po 100 km/h (60 mph) ili da duže trče brzinom od 50 km/h (30 mph). Gazele uglavnom obitavaju u pustinjama, pašnjacima i savanama Afrike; ali se takođe nalaze u jugozapadnoj i centralnoj Aziji i Indijskom potkontinentu. One imaju tendenciju da žive u stadima i jedu manje grube, lako svarljive biljke i lišće. Gazele su relativno male antilope, pri čemu većina ima stojeću visinu od 60-1.070 cm (2-35 ft) u ramenima, i generalno su žućkaste boje.

Slika 26 – Stego tekst

Na strani primaoca je potrebno da se iz stego fajla pročita niz bitova, a potom dekompresuje uz pomoć liste grupa koje određuju prva 2 bita i dodaju se ispred svaka 2 pročitana bita iz teksta. Rezultat je šifrirana poruka koja se pomoću AES dešifratora dešifruje i daje skrivenu poruku.

Ovakav pristup daje visok nivo nevidljivosti, jer se golim okom nikako ne može uočiti prisustvo tajne poruke. Iako se može detektovati prisustvo nevidljivih karaktera nekim programom, ukoliko nije poznat ključ za šifriranje, poruka se ne može izvući. Kompresija dodatno povećeva kapacitet ugradnje, a takođe daje dodatni nivo sigurnosti, jer je za izvlačenje bitova potreban ključ grupa, kao i lista grupa. Međutim, štampanjem se gubi tajna poruka, a takođe se može i obrisati iz word fajla, ukoliko je fajl dostupan da menjanje.

Zaključak

U ovom radu su prikazane metode za sakrivanje teksta u word dokumentima korišćenjem praznog prostora između reči i nevidljivih znakova koji se mogu ubaciti posle svakog karaktera. Implementirani su algoritmi za svaku metodu i date su prednosti i mane kao i njihov kapacitet.

Ne možemo dati tačan i jedinstven odgovor na pitanje koji stego algoritam je najbolji. Istraživači sajber sigurnosti moraju uzeti u obzir mnoge stvari poput različitih prednosti i nedostataka algoritama za skrivanje teksta, zajedno sa preporukama. Takođe bi trebalo razmisliti da li će tehnike skrivanja teksta biti relevantne ili ne za određenu aplikaciju. Treba uočiti koji algoritmi nude koji nivo sigurnosti, nevidljivosti, robusnosti i kapaciteta. Na osnovu toga i potreba korisnika, može se izabrati najpogodniji.

Literatura

[1] – Analiza steganografskih tehnika i metoda - Miroslav Ćajić, Mladen Veinović, Bogdan Brkić,
https://www.researchgate.net/publication/265003223_ANALIZA_STEGANOGRAFSKIH_TEHNIKA_I_METODA

[2] – Steganografija i njene implikacije na forenzičke istrage – Jasmin Ćosić, Miroslav Bača
https://www.researchgate.net/publication/279175018_Steganografija_i_njene_implikacije_na_forenzicke_istrage

[3] – Steganografija kao antiforenzički alat – Manja Đuročkov
https://www.itvestak.org.rs/Zbornik_ZITEH_10/Djurickov%20Manja%20-%20Steganografija.pdf

[4] – A novel aproach for hiding information in text steganography – Wid Akeel
https://www.researchgate.net/publication/324797451_A_Novel_Approach_for_Hiding_Information_in_Text_Steganography

[5] - Modern Text Hiding, Text Steganalysis, and Applications: A Comparative Analysis - Milad Taleby Ahvanooey, Qianmu Li, Jun Hou, Ahmed Raza Rajput, Chen Yini
https://www.mdpi.com/1099-4300/21/4/355

[6] – Techniques for data hiding - W. Bender, D. Gruhl, N. Morimoto, A. Lu
https://pdfs.semanticscholar.org/8c82/c93dfc7d3672e58efd982a23791a8a419053.pdf

[7] - A Novel Approach of Text Steganography based on null spaces - Prem Singh, Rajat Chaudhary, Ambika Agarwal
https://pdfs.semanticscholar.org/1597/51a0c896214a35bb5165bbf5521cd42bd237.pdf

[8] - Steganography in Text by Merge ZWC and Space Character – Ammar Odeh, Khaled Bridgeport
https://www.researchgate.net/publication/256455761_Steganography_in_Text_by_Merge_ZWC_and_Space_Character

[9] – Steganography in text by using MS Word symbols - Ammar Odeh, Khaled Elleithy, Miad Faeizipuor
http://www.asee.org/documents/zones/zone1/2014/Professional/PDFs/35.pdf

[10] - WhiteSteg: A new scheme in information hiding using text steganography – Yee Lip Por, Delina Beh Mei Yin
https://www.researchgate.net/publication/228672143_WhiteSteg_A_new_scheme_in_information_hiding_using_text_steganography?enrichId=rgreq-06d70f693d113018bbc8dae0b6418e79-XXX&enrichSource=Y292ZXJQYWdlOzIyODY3MjE0MztBUzoxMDIyMTg0NTU4NDY5MTJAMTQwMTM4MjE4NzI2Ng%3D%3D&el=1_x_2&_esc=publicationCoverPdf

[11] - A modified approach to data hiding in Microsoft Word documents by change-tracking technique - Susmita Mahato, Danish Ali Khan, Dilip Kumar Yadav

https://www.researchgate.net/publication/319398255_A_modified_approach_to_data_hiding_in_Microsoft_Word_documents_by_change-tracking_technique

[12] – A hybrid text steganography approach utilizing unicode space scharacters and zero-width character – Muhammad Aman, Aihab Khan, Bashir Anhad
https://www.researchgate.net/publication/314449134_A_HYBRID_TEXT_STEGANOGRAPHY_APPROACH_UTILIZING_UNICODE_SPACE_CHARACTERS_AND_ZERO-WIDTH_CHARACTER?enrichId=rgreq-0b9d42f77aee766faa5118644e791c58-XXX&enrichSource=Y292ZXJQYWdlOzMxNDQ0OTEzNDtBUzo0NzAzMDA5NzU0NzI2NDFAMTQ4OTEzOTg5ODU5MA%3D%3D&el=1_x_2&_esc=publicationCoverPdf
[13] - Text steganographic approaches: A comparation - Monika Agarwal
https://www.researchgate.net/publication/235438740_Text_Steganographic_Approaches_A_Comparison