

فهرست مطالب

(۱) مقدمه

(۲) مروری بر روشهای انجام شده

(۳) الگوریتم پیشنهادی استگانوگرافی در متون فارسی

۳ (۱) مروری بر استاندارد یونیکد

۳ (۲) شرح روش پنهان سازی انجام شده

۳ (۳) بررسی و نتیجه گیری

۳ (۴) بررسی الگوریتم روش

مشاور: مؤسسه آموزش عالی روزبهان ساری	مرور و ارائه الگوریتم	کارفرما: وزارت ارتباطات و فناوری اطلاعات
تاریخ آخرین تجدید نظر: ۹۲/۸/۱۴	استگانوگرافی	نسخه: ۱

۱- مقدمه

استگانوگرافی معادل فارسی واژه ی پنهان سازی است که در اصل کلمه ای یونانی بوده و از دو کلمه Steganos به معنای "پنهان کردن" و Graphy به معنای "نوشتن" تشکیل شده است. ترجمه کلمه به کلمه این لغت "نوشته مخفی" است که البته برداشت این معنی از استگانوگرافی چندان متداول نیست و بیشتر به مفهوم پنهان سازی اطلاعات در یک رسانه به عنوان پوشش بکار می رود؛ به گونه ای که توسط اشخاص غیرمجاز قابل تشخیص نباشد.

به طور کلی موضوعاتی که پنهان سازی اطلاعات دربرگیرنده آنها می باشد عبارتند از:

- ۱ - موارد مربوط به حق مالکیت تولیدات نرم افزاری والکترونیکی شامل واترمارکینگ و اثر انگشت که جنبه تجاری از این علم هستند.
- ۲ - استفاده از پنهان سازی در ارسال و دریافت پیام به صورت غیر محسوس

توجه به پنهان سازی اطلاعات از هر دو جنبه بالا دارای اهمیت است. چرا که با فراهم شدن زمینه ه ای فناوری اطلاعات در کشور، لزوم استفاده از قانون حق تکثیر و حفظ حقوق مربوط به مالکیت محصولات نرم افزاری و تولیدات الکترونیکی اعم از موسیقی، آثار هنری، کتاب های الکترونیکی وغیره، شناخت و استفاده از این علم را ایجاب می کند. همچنین پنهان سازی در ترکیب با رمز نگاری قدرت بسیار بالایی را در مقابل حملات مختلف پدید می آورد. شناخت پنهان سازی از جنبه های کنترلی برای پلیس اینترنتی جهت جلوگیری و شناخت معبری برای ارتباطات غیرمجاز و مشکوک نیز دارای اهمیت است.

۲- مروری بر روشهای انجام شده

در این قسمت به بررسی کارهایی که تاکنون بر روی متون انجام گرفته، پرداخته می شود و به دلیل شباهت بسیار زیاد زبان فارسی و عربی روشهای مربوط به این دو زبان بررسی شده است.

۲ ۱ - روش نقطه ها

این روش جزو روشهای مبتنی بر خصوصیات دسته بندی می شود. در الفبای فارسی تعداد زیادی از حروف دارای نقطه هستند. ۱۸ حرف از ۳۲ حرف که ۳ حرف دو نقطه، ۵ حرف سه نقطه و ۱۰ حرف یک نقطه دارند. از بین چهار حرف فارسی که با عربی متفاوت است سه حرف نقطه دار هستند. بنابراین در عربی ۱۵ حرف از ۲۸ حرف نقطه دار هستند. پس می توان گفت تعداد نقطه ها در هر متن فارسی یا عربی قابل توجه است. این روش نیز که بر پایه همین ویژگی الفبای فارسی طراحی شده است توضیح داده می شود.

• روش پنهان سازی

در این روش اطلاعاتی که باید مخفی شوند ابتدا فشرده می شوند. سپس در متن مورد نظر اولین حرف نقطه دار پیدا می شود. با پیدا کردن آن به سراغ اطلاعات فشرده شده می رویم که این اطلاعات به صورت بیتهای ۰ و ۱ هستند. بیت به بیت اطلاعات خوانده می شود اگر مقدار بیت ۰ بود، کاراکتری که از متن برای پنهان سازی انتخاب شده، بدون تغییر باقی می ماند. اما اگر ۱ بود نقطه مربوط به کاراکتر به سمت بالا کمی جا به جا می

مشاور: مؤسسه آموزش عالی روزبهان ساری	مرور و ارائه الگوریتم	کارفرما: وزارت ارتباطات و فناوری اطلاعات
تاریخ آخرین تجدید نظر: ۹۲/۸/۱۴	استگانوگرافی	نسخه: ۱

شود. این روش تکرار شده تا همه اطلاعات در متن مخفی شوند . برای منحرف کردن توجه خوانندگان بعد از مخفی کردن همه اطلاعات، نقاط مربوط به بقیه کاراکترها به صورت تصادفی تغییر می کنند . برای کاراکترهایی که دو یا سه نقطه دارند، همه نقاط تغییر مکان می دهند؛ به دلیل اینکه تغییر دادن یک نقطه از بین نقاط دیگر یک کاراکتر جلب توجه می کند. این روش در شکل ۱-۲ نشان داده شده است.



شکل ۱-۲: جابه جایی عمودی نقطه در روش نقطه ها

• استخراج اطلاعات

برای استخراج اطلاعات، برنامه با شناخت مقدار بیت پنهان شده براساس مکان نقاط روی کاراکترها شروع می شود. بوسیله قرار دادن همه بیت های استخراج شده کنار یکدیگر، اطلاعات فشرده شده بدست می آید. سپس این اطلاعات از حالت فشرده خارج شده و داده های اصلی بدست می آیند

• مزایا

- ۱ - با این روش حجم زیادی از اطلاعات به دلیل وجود تعداد زیاد حروف نقطه دار در فارسی و عربی، می توانند مخفی شوند.
- ۲ - به علت نبود یک برنامه OCR قوی برای زبان های فارسی و عربی، متن چاپی به آسانی به یک متن عادی تبدیل نمی شود؛ بنابراین از بین بردن اطلاعات مخفی شده دشوار است.
- ۳ - متنی که شامل اطلاعات مخفی شده است اختصاص به کامپیوتر ندارد و اطلاعات مخفی می توانند از متن چاپی هم بدست آیند. برای بدست آوردن اطلاعات برای متن چاپی، متن باید اسکن شده و سپس به برنامه مربوطه داده شود.
- ۴ - متن مخفی شده در برابر تغییر سایز فایل مقاوم است و این تغییرات اطلاعات را از بین نمی برد.

• معایب

- ۱- اطلاعات در تایپ دوباره از بین می روند.
- ۲- متن خروجی براساس استفاده از تنها یک فونت قالب ثابتی دارد .(به عبارت دیگر نسبت به تغییرات فونت حساس است)
- ۳ - به دلیل عدم وجود یک برنامه OCR خوب برای زبان های فارسی و عربی، استفاده از این روش در متنی هایی که چاپ شده و سپس دوباره اسکن می شوند مشکل است.

مشاور: مؤسسه آموزش عالی روزبهان ساری	مرور و ارائه الگوریتم	کارفرما: وزارت ارتباطات و فناوری اطلاعات
تاریخ آخرین تجدید نظر: ۹۲/۸/۱۴	استکانوگرافی	نسخه: ۱

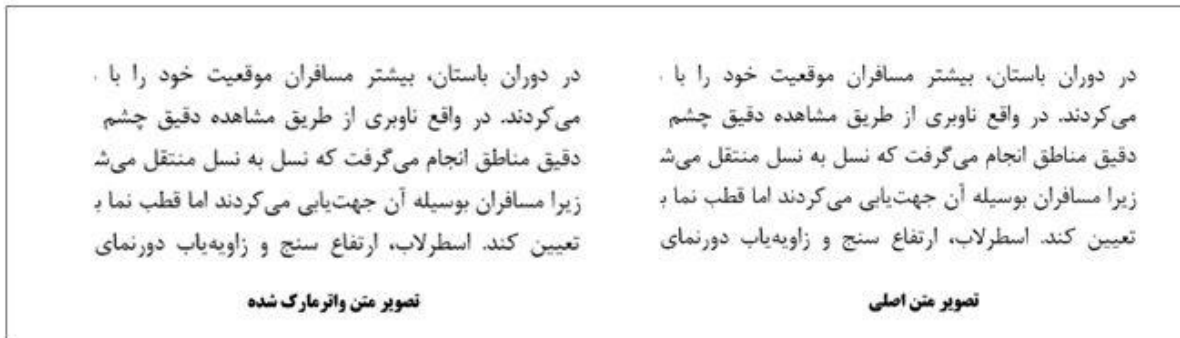
۲ ۴ - استفاده از شکل حروف

در این روش که جزو روشهای واترمارکینگ است و بر روی تصاویر متن کار می کند؛ از چهار حرف (ر ز ژ و) استفاده می شود. این حروف شیب خاصی در شکل شان دارند و می توان از این شیب برای واترمارکینگ داده استفاده کرد. از بین این چهار حرف، سه تای آنها در الفبای عربی هم موجود است بنابراین به جز متون فارسی در متن های عربی نیز می توان از این روش استفاده کرد.

• روش پنهان سازی

سه حرف (ر ز ژ) تنها در تعداد نقاط با یکدیگر متفاوتند بنابراین در این روش پارامتری که برای شیب این سه حرف در نظر گرفته می شود از پارامتر شیب مربوط به حرف (و) جدا می شود. الگوریتم از چهار پارامتر استفاده می کند که دو پارامتر اول مربوط به شیب حروف (ر ز ژ) و حرف (و) در تصویر متن اصلی و دو پارامتر دیگر مربوط به شیب تغییر یافته حروف (ر ز ژ) و حرف (و) در تصویر متن واترمارک شده است.

روش پنهان سازی با استخراج حروف شیب دار (حروف ذکر شده در بالا) از تصویر متن اصلی شروع می شود. اگر بیت مورد نظر برای پنهان سازی ۰ بود تغییری در شیب حرف نمی دهد ولی اگر بیت مورد نظر ۱ بود شیب حرف را تغییر می دهد. این روش ادامه می یابد تا همه اطلاعات مخفی شوند. در شکل ۲-۲ قسمتی از تصویر متن اصلی و تصویر متن واترمارک شده با این روش نشان داده است.



شکل ۲-۲: مقایسه تصویر متن اصلی و تصویر متن واترمارک شده

• استخراج اطلاعات

شیب تمام حروف شیب دار اندازه گیری شده و اختلاف آن با پارامتر مربوط به شیب اصلی حروف مقایسه می شود. اگر اختلافی وجود داشت یعنی بیت پنهان شده در این حرف ۱ است و در غیر اینصورت بیت ۰ می باشد.

• مزایا و معایب

این روش با روشهای جابه جایی خط و کلمه مقایسه شده که مزایایی به شرح زیر از آن مشاهده شده است:

۱- غیر قابل مشاهده بودن (Imperceptibility): نتایج نشان داده است که روش مورد نظر شفاف تر از روشهای جابه‌جایی خط و کلمه است. به دلیل اینکه از بعضی حروف در آن استفاده می‌شود و نه همه آنها، برای سیستم بینایی انسان نامحسوس است.

۲- ظرفیت (Capacity): در مقایسه با بقیه روش‌های معمول ظرفیت بالاتری دارد که در جدول ۲-۲ نشان داده شده است.

۳- کوری (Blindness): این روش می‌تواند جزو روش‌های کور دسته‌بندی شود. به عبارت دیگر رمزگشایی آن به یک کپی از سند اصلی نیاز ندارد تا اطلاعات مخفی شده را استخراج کند. دریافت‌کننده‌های قانونی اسناد واترمارک شده می‌توانند داده‌های واترمارک را تنها با استفاده از پارامترهای شیب بدست آورند. به علاوه استخراج واترمارک به کنترل خط یا کلمه نیاز ندارد که این امر باعث مصون شدن در برابر خطاهایی می‌شود که در قطعه بندی صفحات رخ می‌دهد.

• بررسی

ظرفیت این روش با روش‌های جابه‌جایی کلمه و خط در یک متن نمونه فارسی مقایسه شده و نتایج آن در جدول ۱-۲ آورده شده است.

جدول ۱-۲: مقایسه ظرفیت در روش‌های واترمارکینگ برای یک متن نمونه فارسی

نام روش	ظرفیت پنهان سازی (برحسب بیت)
جابه‌جایی خط	۱۵
جابه‌جایی کلمه	۲۲۲
روش مورد نظر	۳۳۵

دو روش شرح داده شده برای پنهان سازی از روش در تغییر ظاهر متن استفاده می‌کردند اما روش‌هایی که در ادامه توضیح داده می‌شود از استاندارد یونیکد برای عمل پنهان سازی استفاده می‌کنند

۲ ۴- استفاده از کاراکتر کشش

در استاندارد یونیکد کاراکتری با کد شانزدهی ۰۶۴۰ موجود است. این کاراکتر به عنوان یک کاراکتر مازاد تنها برای اهداف ساختاری و چیدمانی استفاده می‌شود. این کاراکتر در همه جا قابل استفاده نیست و تنها در فضای بین حروف متصل شونده عربی استفاده می‌شود به عبارت دیگر بعد از حروف آخر کلمات یا قبل از حروف اول کلمات نمی‌آید. وقتی این کاراکتر بین دو حرف متصل شونده قرار می‌گیرد اندکی بین دو حرف کشش ایجاد می‌کند. از این کاراکتر برای اهداف پنهان سازی استفاده شده است که در ادامه توضیح می‌دهیم.

• روش پنهان سازی

برای پنهان کردن یک رشته بیتی به عنوان مثال با مقدار "۱۱۰۰۱۰" از کم ارزش ترین بیت که در این مثال ۰ است شروع می کند. اگر بیت مربوطه ۰ بود، در اولین حرف بدون نقطه ای که بعد از آن امکان اضافه کردن کاراکتر کشش وجود داشته باشد؛ بیت را با اضافه کردن کاراکتر کشش مخفی می کند. اگر بیت مربوطه ۱ بود، در اولین حرف نقطه داری که بعد از آن امکان اضافه کردن کاراکتر کشش وجود داشته باشد بیت را مخفی می کند. به عبارت دیگر بیت ۰ را با اضافه کردن کاراکتر کشش بعد از حروف بدون نقطه و بیت ۱ را با اضافه کردن کاراکتر کشش بعد از حروف نقطه دار پنهان سازی می کند. این روش در جدول ۲-۲ نشان داده شده است.

جدول ۲-۲: روش پنهان سازی با اضافه کردن کاراکتر کشش بعد از حروف

Watermarking bits	۱۱۰۰۱۰
Cover-text	من حسن اسلام المرء ترکه مالا یعنی
Output text	من حسن اسلام المرء ترکه مالا یعنی ↑ ↑ ↑ ↑ ↑ ↑
	۱ ۱ ۰ ۰ ۱ ۰

به علاوه می توان قبل از حروف نیز کاراکتر کشش را اضافه کرد. این روش نیز در جدول ۳-۲ نشان داده شده است.

جدول ۳-۲: روش پنهان سازی با اضافه کردن کاراکتر کشش قبل از حروف

Watermarking bits	۱۱۰۰۱۰
Cover-text	من حسن اسلام المرء ترکه مالا یعنی
Output text	من حسن اسلام المرء ترکه مالا یعنی ↑ ↑ ↑ ↑ ↑ ↑
	۱ ۱ ۰ ۰ ۱ ۰

برای افزایش امنیت و گمراه کردن متجاوزان، در یک متن ثابت در خطها و پاراگرافهای متفاوت می توان از هر دو روش گفته شده در بالا استفاده کرد.

• استخراج اطلاعات

استخراج اطلاعات در این روش عکس عمل پنهان سازی است. اگر بعد(قبل) از حرف نقطه دار کاراکتر کشش وجود داشت، بیت پنهان شده ۱ است و اگر بعد(قبل) از حرف بدون نقطه کاراکتر کشش بود بیت پنهان شده ۰ است

• نتایج

این روش هر سه خصوصیت ظرفیت، امنیت و قوت که جنبه های لازم برای پنهان سازی هستند را به خوبی بر آورده می کند. به علاوه به دلیل استفاده از قابلیت های یونیکد در محتوای نوشتاری نیز تغییری ایجاد نمی کند. اما برای پنهان سازی هر بیت، یک کاراکتر به متن اضافه می کند و به این ترتیب حجم فایل افزایش می یابد.

ظرفیت این روش برای پنهان سازی انواع مختلف فایلها در جدول ۲-۴ نشان داده شده است.

جدول ۲-۴: مقایسه ظرفیت روش استفاده از کاراکتر کشش در نوع های مختلف فایل

نوع فایل	سایز فایل (بایت)	سایز فایل پوششی (بایت)	ظرفیت (%)
.txt	۴۴۳۹	۳۶۵۱۸۱	۱.۲۱۵
.html	۴۴۳۹	۳۷۸۵۸۹	۱.۱۷۲
.cpp	۱۰۱۲۷	۷۹۹۵۷۷	۱.۲۶۶
.gif	۱۸۸	۱۵۱۱۲	۱.۲۴۴
میانگین		۱.۲۲	

۲ ۴ - روش توسعه یافته استفاده از کاراکتر کشش

ایده اصلی این روش افزایش دادن ظرفیت پنهان سازی با استفاده از همه موقعیت های ممکن برای استفاده از کاراکتر کشش در حروف عربی است. ۲۸ حرف در زبان عربی وجود دارد که بعضی از این حروف بیشتر از یک شکل دارند برای مثال حرف "ش" شکل متفاوت دارد {ا | اؤ | اء | ائ}. به این صورت ۲۸ حرف عربی می توانند با ۳۵ شکل مختلف ظاهر شوند. در این روش تمام مکان های بین حروف که امکان اضافه کردن کاراکتر کشش قبل یا بعد از آنها وجود دارد با استفاده از تمام شکل های حروف عربی بررسی شده است. این روش که در واقع توسعه یافته روش قبلی است پنهان سازی و استخراج پیام در آن نیز مانند قبلی می باشد. این روش MSCUKAT نامیده شده است.

• نتایج

مقایسه این روش با روش استفاده از کاراکتر کشش نشان می دهد که پیشرفت قابل توجهی در ظرفیت پنهان سازی در حدود ۱۸۶ درصد مشاهده شده است.

در یک آزمایش مقایسه ای حجم پیام رمز به طول ۳۵۲ بیت ثابت در نظر گرفته شده و حجم متن پوششی تغییر داده شده و عمل پنهان سازی در روش مورد نظر و روش استفاده از کاراکتر کشش با هم مقایسه شده است که قسمتی از نتایج آن در جدول ۲-۵ نشان داده شده است.

جدول ۲-۵: مقایسه ظرفیت روش MSCUKAT با روش استفاده از کاراکتر کشش در حالت طول پیام رمز ثابت

ردیف	طول رسانه پوششی	روش MSCUKAT		روش استفاده از کاراکتر کشش	
		ظرفیت	درصد	ظرفیت	درصد
۱	۲۳۵۷	۸۶۱	۴۰.۸۸	۱۶۵۳	۲۱.۲۹

۱۹.۷۲	۱۷۸۵	۴۱.۶۶	۸۴۵	۲۵۰۳	۲
۲۱.۳۵	۱۶۴۹	۳۶.۰۳	۹۷۷	۲۹۰۵	۳
۲۰.۲۲	۱۷۴۱	۳۸.۷۲	۹۰۹	۲۹۹۰	۴
۲۱.۰۰		۳۹.۰۰		میانگین	

آزمایش دیگری که برای مقایسه این دو روش انجام شده طول متن پوششی با ۵۵۶۷ کاراکتر را ثابت و طول پیام رمز را تغییر داده است. قسمتی از نتایج آن در جدول ۲-۶ نشان داده شده است.

جدول ۲-۶: مقایسه ظرفیت روش MSCUKAT با روش استفاده از کاراکتر کشش در حالت طول متن ثابت

روش استفاده از کاراکتر کشش		روش MSCUKAT		طول پیام رمز	ردیف
درصد	ظرفیت	درصد	ظرفیت		
۲۰.۸۸	۹۹۶	۳۸.۵۲	۵۴۰	۲۰۸	۱
۲۰.۵۵	۱۰۹۰	۳۸.۶۲	۵۸۰	۲۲۴	۲
۲۱.۶۹	۱۶۲۳	۴۰.۰۰	۸۸۰	۳۵۲	۳
۲۰.۰۷	۱۶۲۳	۳۹.۷۲	۸۴۶	۳۳۶	۴
۲۰.۹۰		۳۹.۵۰		میانگین	

هر دو آزمایش بالا نشان دهنده اینست که روش MSCUKAT ظرفیت بالاتری نسبت به روش استفاده از کاراکتر کشش دارد

۲ ۵ - روش پنهان سازی با استفاده از قابلیت‌های یونیکد

در استاندارد یونیکد، برای حروف عربی دو حالت ذخیره سازی کد وجود دارد یک کد که نماینده حرف است و بستگی به موقعیت حرف در کلمه ندارد. نوع دیگر ذخیره سازی کد بستگی به موقعیت حرف در کلمه دارد و به چهار نوع ظاهر می‌شود و برای هر هر چهار نوع یک کد جدا وجود دارد.

در یونیکد تنها کد نماینده حرف در فایل متنی ذخیره می‌شود و برنامه بر حسب موقعیت حرف در کلمه شکل صحیح را نشان می‌دهد. اگرچه می‌توان کد شکلی را نیز ذخیره کرد. از این ویژگی یونیکد در این روش استفاده شده است.

• روش پنهان سازی

برای پنهان سازی بیت ۰، کد نماینده حروف در هر کلمه ذخیره می‌شود و برای بیت ۱، کد شکلی حروف در کلمه ذخیره می‌شود. نمی‌توانیم از کد شکلی و کد نماینده حروف در یک کلمه با یکدیگر استفاده کنیم به دلیل اینکه برنامه نمی‌تواند شکل حروف را در کلمه به طور صحیح تشخیص دهد.

• استخراج اطلاعات

مشاور: مؤسسه آموزش عالی روزبهان ساری	مرور و ارائه الگوریتم	کارفرما: وزارت ارتباطات و فناوری اطلاعات
تاریخ آخرین تجدید نظر: ۹۲/۸/۱۴	استکانوگرافی	نسخه: ۱

کد حروف کلمه چک می شود. اگر کدهای ذخیره شده کلمه، کد نماینده حروف باشد نتیجه می گیریم کد ۰ مخفی شده اما اگر با شکل حرفی ذخیره شده باشد کد ۱ مخفی شده است.

• نتایج

ظرفیت ذخیره سازی این روش تقریباً با روش نقطه ها برابر است. این روش هر سه پارامتر مهم در طراحی روش های پنهان سازی را برآورده می کند. ظرفیت خوبی دارد و یک بیت را در هر کلمه مخفی می کند

۲ ۶ - روش بهبود یافته "لا"

کلمه "لا" یک کاراکتر با کد FEFB در استاندارد یونیکد است. می توان به صورت دیگری نیز کلمه لا را نشان داد که به صورت (حرف ل+کاراکتر فاصله بین حروف با کد شانزدهی ۰۶۴۰+حرف ا) است که به صورت "لا" نشان داده می شود.

• روش پنهان سازی

برای پنهان سازی بیت ۱ از کاراکتر "لا" و برای بیت ۰ از "لا" استفاده می شود. این روش در نمونه قبلی لا استفاده می شده است اما به دلیل مشکلاتی که این روش داشته است از قبیل افزایش حجم فایل و به دلیل گذاشتن فاصله بین "ل" و "ا" کلمه را پهن تر کرده و ظاهر متن را غیرطبیعی می کند.

در روش بهبود یافته به جای وارد کردن کد فاصله بین "ل" و "ا" کد شکلی آنها آورده می شود و کلمه به صورت "لا" نشان داده می شود.

• استخراج اطلاعات

عکس عمل پنهان سازی است و اگر در متن "لا" باشد بیت مخفی شده ۱ و اگر "لا" باشد بیت مخفی شده ۰ است.

• نتایج

این روش مشکلات روش قبلی لا را برطرف کرده است.

۲ ۷ - استفاده از حروف مشابه با کدهای متفاوت

• روش پنهان سازی

بین بعضی از حروف عربی و فارسی مثل "ی" و "ک" در شکل ظاهری آنها تفاوت وجود دارد برای مثال ی فارسی به صورت "ی" و ی عربی به صورت "ی" و حرف ک نیز در فارسی بصورت "ک" و در عربی به صورت "ک" نشان داده می شود.

مشاور: مؤسسه آموزش عالی روزبهان ساری	مرور و ارائه الگوریتم	کارفرما: وزارت ارتباطات و فناوری اطلاعات
تاریخ آخرین تجدید نظر: ۹۲/۸/۱۴	استگانوگرافی	نسخه: ۱

از این قابلیت می توان برای پنهان سازی استفاده کرد . این روش به این صورت است : برای پنهان کردن بیت ۱ از شکل عربی حروف و برای بیت ۰ از شکل فارسی آنها استفاده می شود.

• نتایج

این روش شفافیت بالایی دارد ولی ظرفیت پنهان سازی آن خیلی زیاد نیست . در مقابل بعضی از حملات مثل تایپ مجدد آسیب پذیر است ولی در مقابل copy-paste آسیب پذیر نیست.

۲ A - استفاده از فاصله نمادین بین کلمات

در فارسی و عربی، بعلاوه فاصله معمولی که بین کلمات وجود دارد . در بعضی کلمات مثل "رفته ام" یک فاصله کم بین دو قسمت یک کلمه وجود دارد . این فاصله به عنوان فاصله نمادین شناخته می شود. در سیستم یونیکد فاصله معمولی کد شانزدهمی ۰۰۲۰ را دارد ولی فاصله نمادین که کد ZWNJ نیز نامیده می شود کد ۲۰۰C را دارد.

• روش پنهان سازی

مخفی کردن اطلاعات در این روش به این صورت است که اگر در کلمه ای فاصله نمادین وجود داشت ، ما یکی از دو فرم فاصله را بسته به اطلاعاتی که می خواهیم در متن پنهان کنیم انتخاب می کنیم. برای پنهان کردن بیت ۱ فاصله معمولی را بعد از فاصله نمادین در کلمه اضافه می کنیم و اگر بخواهیم بیت ۰ را پنهان کنیم ، فاصله نمادین را تغییر نمی دهیم.

• استخراج اطلاعات

اگر در کلمه ای از فاصله نمادین استفاده شده بود یعنی بیت ۰ در آن پنهان شده و اگر هم از فاصله نمادین و هم فاصله معمولی بیت ۱ در آن پنهان شده است.

• نتایج

ظرفیت پنهان سازی این روش در حدود ۴ بیت در هر کیلوبایت است. این روش وابسته به هیچ فونت و فرمتی نیست. در فرمت های بسیاری مانند HTML یا Word قابل استفاده است. اما سایز متن را خیلی زیاد می کند.

۲ B - استفاده از عبارات منظم ZWJ و ZWNJ

در سیستم یونیکد کاراکتر ZWJ با کد ۲۰۰D و کاراکتر ZWNJ با کد ۲۰۰C جز کاراکترهای شفاف هستند. کاراکتر ZWJ وقتی بین دو حرف قرار می گیرد باعث اتصال دو حرف به یکدیگر می شود و کاراکتر ZWNJ برخلاف آن هنگام قرار گرفتن بین دو حرف آن دو را از یکدیگر جدا می کند. برای مثال "مZWNJح" به صورت "م ح" و "مZWJح" به صورت "م ح" دیده می شود.

• روش پنهان سازی

مشاور: مؤسسه آموزش عالی روزبهان ساری	مرور و ارائه الگوریتم	کارفرما: وزارت ارتباطات و فناوری اطلاعات
تاریخ آخرین تجدید نظر: ۹۲/۸/۱۴	استکانوگرافی	نسخه: ۱

این روش با بهبود روش ۴-۱-۶ اطلاعات را به جای پنهان کردن یک بیت درون حرفها، درون کلمات پنهان می کند. در این حالت کاراکتر ZWj معادل بیت ۱ و ZWnj معادل بیت ۰ هستند. برای پنهان سازی، یک گروه از ZWj و ZWnj ها با سایز بلوک مشخص به کلمات اضافه می شود. از دو نوع عبارت منظم برای پنهان سازی استفاده می کند.

۱- عبارات منظم ZWnj: این عبارات به این صورت می آیند:

(یک حرف غیر متصل یا کاراکتر فاصله +ZWNj+ گروهی از ZWj یا ZWNjها+ یک حرف غیر متصل یا کاراکتر فاصله)

که حروف غیر متصل در زبان عربی شامل (د، ذ، ر، ز، و، ا) هستند. جدول ۲-۷ چگونگی مخفی کردن بیت ها را به این روش نشان می دهد.

جدول ۲-۷: مخفی کردن بیت ها با بلوک ZWNj

Embedded Bits	Code Points	Appearance
۱۰۱۱۱۱۰۰ ۰۰۱۱۰۱۱۱	+۲۰۰C+۲۰۰D+۲۰۰C+۲۰۰D+۲۰۰D+۲۰۰D+ +۲۰۰D+۲۰۰C+۲۰۰C+۲۰۰C+۲۰۰C+۲۰۰C+۲۰۰D+۲۰۰D+ +۲۰۰C+۲۰۰D+۲۰۰D+۲۰۰D+ _r	ور

۲- عبارات منظم ZWj: این عبارات به این صورت می آیند:

(یک حرف متصل شونده از سمت چپ یا متصل شونده از هر دو طرف + ZWj + گروهی از ZWj یا ZWNj ها + یک متصل شونده از سمت راست یا متصل شونده از هر دو طرف)

جدول ۲-۸ چگونگی مخفی کردن بیت ها در این روش را نشان می دهد.

جدول ۲-۸: مخفی کردن بیت ها با بلوک ZWj

Embedded Bits	Code Points	Appearance
۱۰۱۱۱۱۰۰ ۰۰۱۱۰۱۱۱	+۲۰۰D+۲۰۰D+۲۰۰C+۲۰۰D+۲۰۰D+۲۰۰D+ _m +۲۰۰C+۲۰۰C+۲۰۰C+۲۰۰C+۲۰۰D+۲۰۰D+۲۰۰C+ +۲۰۰D+۲۰۰D+ _h	مح

• نتایج

این روش از متن های عربی به عنوان متن پوششی استفاده کرده و فایل های متنی را با سایز بلوک مشخص در آن مخفی نموده است که نتایج آن در جدول ۲-۹ آمده است.

مشاور: مؤسسه آموزش عالی روزبهان ساری	مرور و ارائه الگوریتم استگنوگرافی	کارفرما: وزارت ارتباطات و فناوری اطلاعات
تاریخ آخرین تجدید نظر: ۹۲/۸/۱۴		نسخه: ۱

با توجه به اینکه میانگین ظرفیت روش نقطه هادر حدود ۱.۳۷ و میانگین روش توسعه یافته استفاده از کاراکتر کششدر حدود ۳.۰۱۹ است. این روش نسبت به روش های قبلی میانگین ظرفیت بالایی در حدود ۱۳ درصد را دارا می باشد. همچنین به فرمت متنی خاصی نیاز دارد و میزان شفافیت آن نیز زیاد است.

جدول ۲-۹: مقایسه ظرفیت روش استفاده از عبارات منظم

میانگین ظرفیت (%)	ظرفیت (%)	سایز بلوک مخفی شده	سایز متن پوششی (بایت)
۱۳.۲۵۵	۳.۵	۸	۴۹۲۴
	۷.۰۷	۱۶	
	۱۴.۰۶	۳۲	
	۲۸.۳۹	۶۴	
۱۳.۰۷	۳.۴۷	۸	۱۸۵۶۳
	۶.۹۶	۱۶	
	۱۳.۹۵	۳۲	
	۲۷.۹۰	۶۴	

۳- الگوریتم پیشنهادی استگنوگرافی در متون فارسی

در این بخش روش پیشنهادی را اجرا شده است را بررسی نموده و مراحل و اجزای مختلف آن را توضیح می دهیم. با توجه به اینکه در این روش از قابلیت های یونیکد استفاده شده است ابتدا به توضیح این استاندارد می پردازیم.

۳-۱- مروری بر استاندارد یونیکد

یونیکد یکی از استانداردهایی است که برای انکدینگ کاراکترها استفاده می شود. این استاندارد با بیشتر استانداردهای انکدینگ مانند ISO/IEC ۱۰۶۴۶ هماهنگ است.

۳-۱-۱- تعیین کدهای کاراکترها

یک شماره به هر کاراکتر یا جز کد اختصاص داده شده است. هر کدام از این عدد ها یک کدپوینت هستند. هنگام مراجعه به متن در فرم شانزدهی با پیشوند "U+" می آیند. هر کاراکتر یک نام یکتا نیز دارد برای مثال کاراکتر انگلیسی "A" با کد دهدهی ۶۵ در سیستم یونیکد به صورت "U+۰۰۴۱" نشان داده می شود. این نامهای یونیکد برای کاراکترهای یکسان در ISO/IEC ۱۰۶۴۶ هماهنگ هستند.

استاندارد یونیکد کاراکترها را بوسیله اسکریپت هایی در بلوک ها گروه بندی می کند. یک اسکریپت یک سیستم از کاراکترهای مربوط است. استاندارد، کاراکترها را در یک مجموعه منبع نگه می دارد.

مشاور: مؤسسه آموزش عالی روزبهان ساری	مرور و ارائه الگوریتم	کارفرما: وزارت ارتباطات و فناوری اطلاعات
تاریخ آخرین تجدید نظر: ۹۲/۸/۱۴	استکانوگرافی	نسخه: ۱

زمانی که کاراکترهای یک زبان در یک ترتیب خاص هستند مثلا ترتیب الفبایی، استاندارد یونیکد هم آنها را در ناحیه کدشان طوری مشخص می کند که از همان ترتیب استفاده می کنند. بلوک ها در سایز متفاوتند مثلا بلوک Cyrillic، ۲۶۵ کدپوینت دارد ولی CJK شامل هزاران کدپوینت است.

هر زبان در یک بلوک به ترتیب الفبایی مرتب شده است. کدینگ با "U+۰۰۰۰" شروع شده و با کاراکترهای اسکی و بعد یونانی، سریلانکایی، عبری، عربی، هندی و بقیه زبان ها و بعد علائم و نمادهای نقطه گذاری ادامه پیدا می کند و همینطور بقیه موارد موجود در یونیکد می آید. یک محدوده از کدپوینت ها برای استفاده خصوصی رزرو شده اند که در ادامه یک محدوده از کاراکترهای سازگاری می آیند. این کاراکترها برای تطبیق با نسخه های قدیمی ساخته شده اند. محدوده وسیعی از کدپوینت ها روی BMP و دو محدوده خیلی وسیع در سطح تکمیلی به عنوان نواحی استفاده خصوصی رزرو شده اند. که هیچ معنای عمومی ندارند و برای استفاده کاربران یا برنامه ها هستند. مجموعه ای از برنامه های صفحه بندی ممکن است از آنها به عنوان کدهای کنترلی برای قرارگیری متن در صفحه استفاده کند.

۳-۱-۲- فرمهای انکدین

استاندارد های انکدینگ نه تنها برای هر کاراکتر مقدار عددی یا کدپوینت آن را تشخیص می دهند بلکه چگونگی نمایش مقدار در بیت ها را هم نشان می دهند.

این استانداردها سه نوع فرم انکدینگ را تعریف می کنند که اجازه می دهد داده یکسان در سه حالت فرستاده شود. بایت، کلمه و فرمت دابل کلمه گرا به عبارت دیگر ۸، ۱۶ و ۳۲ بیت برای هر واحد کد در نظر گرفته می شود. هر سه از همان مجموعه کد رایج استفاده می کنند و می توانند بدون از دست دادن داده به یکدیگر تبدیل شوند. این سه فرم توضیح داده شده است:

۸-UTF: برای HTML و پروتکل های مشابه متداول است. راهی برای تبدیل همه کاراکترها به یک مقدار از بایت ها مانند بایت های ASCII دارد.

۱۶-UTF: در محیطی متداول است که نیاز به تعادل بین استفاده موثر از کاراکترها با استفاده اقتصادی از حافظه است. فشرده است و همه کاراکترهای مورد استفا ده سنگین در یک واحد کد ۱۶ بیتی هستند در حالی که بقیه کاراکترها هم از طریق جفت واحدهای کد ۱۶ بیتی قابل دسترسی اند.

۳۲-UTF: در جایی مفید است که فضای حافظه مهم نباشد اما پهنای ثابت دسترسی به واحد کد مطلوب باشد. هر کاراکتر یونیکد در یک کد تک ۳۲ بیتی انکد می شود.

همه فرمهای انکدینگ حداکثر به ۴ بایت یا ۳۲ بیت از داده برای هر کاراکتر نیاز دارند.

۳-۱-۳- یونیکد فارسی

با توجه به شباهت زبان فارسی با عربی، کاراکترهای فارسی در بلوک مربوط به زبان عربی هستند. بلوک عربی از ۰۶۰۰ شروع شده و تا ۰۶FF ادامه می یابد. ۰۶ اولی که در کاراکترهای عربی یا فارسی می آید نشان دهنده اینست که

مشاور: مؤسسه آموزش عالی روزبهان ساری	مرور و ارائه الگوریتم	کارفرما: وزارت ارتباطات و فناوری اطلاعات
تاریخ آخرین تجدید نظر: ۹۲/۸/۱۴	استگنوگرافی	نسخه: ۱

این کاراکتر متعلق به بلوک عربی است و دو رقم بعدی کد خود کاراکتر است . به عنوان مثال حرف "م" با کد ۰۶۴۵ شناخته می‌شود.

۲-۳ - شرح روش پنهان سازی انجام شده

ما در این روش، عمل پنهان سازی را با مخفی کردن دو رقم دودویی در یک حرف انجام می‌دهیم. برخلاف بسیاری از روش‌های دیگر که در هر بار عمل پنهان سازی تنها یک رقم مخفی می‌شود در این روش در هر بار اجرای الگوریتم، دو رقم مخفی می‌شود که ظرفیت روش را بالاتر می‌برد. همچنین در این روش داده در تمام حروف پنهان می‌شود نه در حرف خاصی، بنابراین محدودیت کمتر می‌شود. در ادامه اجزای روش موردنظر شرح داده شده است.

عمل پنهان سازی ارائه شده در چند مرحله انجام می‌گیرد:

۳-۲-۱- تبدیل پیام رمز به معادل دودویی

ابتدا پیام رمز برای مخفی شدن درون متن باید به حالت دودویی تبدیل شود . در این روش، تبدیل به معادل دودویی با روش‌های دیگر پنهان سازی متفاوت است و الگوریتم خاصی برای آن در نظر گرفته شده است.

شرح روش:

با توجه به اینکه ۳۲ حرف در الفبای فارسی وجود دارد و کاراکتر فاصله نیز در پیام رمز استفاده می‌شود. ما برای هر حرف ، ۶ رقم دودویی در نظر می‌گیریم. واضح است که با ۶ رقم دودویی، ۲^۶ یعنی ۶۴ حرف را می‌شود کدبندی کرد. ما در این‌جا، تنها حروف فارسی و کاراکتر فاصله را کدبندی کرده ایم ولی می‌توان این کار را تا ۶۴ کاراکتر نیز گسترش داد. در حروف الفبای فارسی حرف "ا" معمولا به دو شکل "آ" و "ا" ظاهر می‌شود. به همین دلیل ما حروف الفبا را از "آ" شروع کرده و سپس بقیه حروف به ترتیب الفبا و در آخر نیز کاراکتر فاصله می‌آید. این کاراکترها در یک آرایه ذخیره شده و به هرکدام از آنها، اعداد ۰ تا ۳۳ را نسبت می‌دهیم. معادل دودویی هر کاراکتر در این روش در واقع همان عدد ترتیب نسبت داده شده است که به مبنای دو آورده می‌شود.

به طور خلاصه الگوریتم تبدیل پیام رمز به معادل دودویی شامل مراحل زیر است:

- ۱- پیام رمز وارد می‌شود (حتما باید از حروف فارسی و کاراکتر فاصله تشکیل شده باشد)
- ۲- پیام رمز کاراکتر به کاراکتر خوانده شده و هر کاراکتر با آرایه ای که در آن حروف الفبا ذخیره شده مقایسه می‌گردد تا معلوم شود در ترتیب الفبایی چه عددی دارد.
- ۳- عدد ترتیب به دست آمده به یک عدد ۶ رقمی در مبنای ۲ تبدیل می‌شود.
- ۴- این روش ادامه می‌یابد تا کل پیام رمز به معادل دودویی تبدیل شود.

به عنوان مثال کلمه "در" دارای معادل دودویی با مقدار "۰۰۱۰۱۰۰۰۱۱۰۰" است. که ۶ رقم اول معادل حرف "د" هستند و اگر به حالت دهدهی تبدیل شود مقدار آن ۱۰ است. این عدد شماره ترتیبی است که در آرایه حروف الفبا به آن داده شده است. ۶ رقم بعدی هم معادل حرف "ر" است.

۳-۲-۲- شرح روش پنهان سازی

با توجه به اینکه در هر بار عمل پنهان سازی دو رقم دودویی مخفی می شود ما ارقام دودویی را به چهار گروه تقسیم می کنیم: (۱۱۰، ۱۰۱، ۰۱۰ و ۰۱۱)

به دلیل اینکه ارقام درون حروف مخفی می شوند. بررسی ای را بر روی معادل دودویی کار اکثرها که از مرحله قبل بدست می آید انجام دادیم. در بررسی انجام شده روی تمام معادل های دودویی مشخص شد که تعداد "۰۰"ها ۳۵، "۰۱"ها ۳۳، "۱۰"ها ۱۸ و "۱۱"ها ۱۶ تا هستند.

بنابراین تعداد "۰۰"ها از همه بیشتر است. ما بررسی بر روی تعداد تکرار حروف فارسی در متون انجام دادیم و حروفی که به نسبت درصد تکرار بیشتری داشتند را برای جاسازی "۰۰"ها قرار دادیم همچنین بقیه گروه های حروف نیز بر همین اساس مرتب شدند.

۳۲ حرف در الفبای فارسی وجود دارد و حرف "ا" نیز دو شکل "آ" و "ا" دارد. چون چهار حالت مختلف برای پنهان سازی داریم: (۱۱۰، ۱۰۱، ۰۱۰ و ۰۱۱). حروف الفبا را نیز به چهار گروه تقسیم می کنیم و در هر گروه از حروف دو رقم دودویی مخفی می شوند. برای افزایش ظرفیت علاوه بر حروف الفبا ما از کاراکتر های فاصله و کشش و دو کد "zwj" و "zwnj" نیز استفاده کردیم. به دلیل اینکه گروه "۰۰" مربوط به حروف جدا هستند و کد "zwnj" نیز برای جدا کردن حروف به کار می رود، در این گروه قرار گرفته است. ولی کد "zwj" که برای اتصال حروف استفاده می شود، تفاوتی ندارد که در کدامیک از گروههایی که حروف متصل دارند قرار گیرد. ما بر حسب موقعیت گروه ها این کد را در گروه "۱۱" قرار دادیم. در بررسی که بر روی کاراکتر کشش با کد شانزدهمی ۰۶۴۰ انجام گرفت، این مطلب مشاهده شد که هر کدام از کدهای نشانه که در قسمت بعدی توضیح داده شده، بعد از آن بیاید تغییری در ظاهر متن ایجاد نمی شود. بنابراین ما این کاراکتر را در گروه "۰۱" قرار دادیم.

این گروه ها در جدول ۲-۱۰ نشان داده شده است.

جدول ۲-۱۰: گروه های حروف برای پنهان سازی داده

نام گروه	کاراکترهای موجود در گروه
۰۰	آ ا د ز ر ژ و فاصله "zwnj"
۰۱	س ک گ ل م ن ه ی "کاراکتر کشش"
۱۰	پ ت ث ج چ ع غ ف ق
۱۱	ب ش ح خ ص ض ط ظ "zwj"

داده ای که در هر حرف موجود در هر گروه، مخفی می شود با نام آن گروه یکسان است.

۳-۲-۳ - شیوه نشانه گذاری برای حروف

برای اینکه مشخص کنیم در حرف مورد نظر ما داده مخفی شده است باید نشانه ای را برای آن حرف تعیین کنیم. این روش در ادامه شرح داده می شود:

با توجه به اینکه یکی از ویژگی های پنهان سازی، شفافیت و این موضوع است که متن استگو با متن اصلی نباید تفاوت محسوسی داشته باشد. به علاوه در این روش ما می خواهیم داده ها را درون حروف مخفی کنیم از قابلیت های یونیکد برای نشانه گذاری حروف استفاده می کنیم.

همانطور که در مقاله های [۱] و [۲] اشاره شد. دو کاراکتر در یونیکد با نامهای "zwj" و "zwnj" وجود دارد که یکی برای اتصال حروف به یکدیگر و دیگری برای جدا کردن حروف به کار می رود. ما از این کاراکترها برای نشانه گذاری استفاده می کنیم این کار به این صورت انجام می گیرد:

اگر حرف مورد نظر ما در گروه "۰۰" بود یعنی این حرف جز حروف جدا است و به حرف بعد از خود نمی چسبد بنابراین تفاوتی ندارد که حرف بعد از آن چه باشد و ما به همین دلیل شرطی برای آن در نظر نمی گیریم. برای نشانه گذاری حروف موجود در این گروه بعد از آن ها دو کاراکتر "zwnj" با کد شانزدهی C۲۰۰ وارد می کنیم.

اگر حرف ما در گروهی غیر از گروه "۰۰" بود چون این حروف به حرف بعد از خود می چسبند باید حرف بعد از آن را بررسی کنیم. اگر حرف بعد از آن جز حروف الفبا بود یعنی اینکه بین دو حرف اتصال وجود دارد و برای نشانه گذاری از دو کاراکتر "zwj" با کد شانزدهی D۲۰۰ وارد می کنیم. چون این کد برای اتصال حروف است و حرف مورد نظر ما هم به حرف بعد از خود چسبیده است وارد کردن این دو کد تغییری در ظاهر آن ایجاد نمی کند. اما اگر کاراکتر بعد از آن غیر از حروف الفبا باشد از دو کد "zwnj" استفاده می کنیم که این کار هم تغییری در ظاهر متن ایجاد نمی کند. مثلا اگر حرف مورد نظر ما "ب" باشد و کاراکتر بعدی آن علامت "(" باشد اضافه کردن این دو کد هیچ تغییری در ظاهر بوجود نمی آورد. انجام این کار باعث بالا بردن ظرفیت پنهان سازی می شود ما در هر صورت می توانیم داده را در هر حرف مخفی کنیم.

اما یک استثنا در این مورد وجود دارد اگر بعد از حرف مورد نظر کاراکتر کشش با کد ۰۶۴۰ باشد ما آن حرف را به عنوان نشانه در نظر نمی گیریم. با توجه به بررسی انجام شده اگر بین دو کاراکتری که یک حرف الفبایی و دیگری کاراکتر کشش است، هر کدام از کدهای "zwj" یا "zwnj" قرار گیرند ظاهر متن به هم می ریزد و حالت غیر طبیعی پیدا می کند.

به این دلیل به جای یک کد از دو کد "zwj" یا "zwnj" استفاده می کنیم چون ممکن است در خود متن اصلی این کاراکترها وجود داشته باشند و با نشانه مورد نظر ما اشتباه گرفته شوند. اضافه کردن دو کاراکتر روش را مطمئن تر می کند.

۳-۲-۴ - الگوریتم پنهان سازی

با توجه به موارد گفته شده می توان الگوریتم پنهان سازی را به طور خلاصه به شرح زیر بیان نمود:

مشاور: مؤسسه آموزش عالی روزبهان ساری	مرور و ارائه الگوریتم	کارفرما: وزارت ارتباطات و فناوری اطلاعات
تاریخ آخرین تجدید نظر: ۹۲/۸/۱۴	استگانوگرافی	نسخه: ۱

۱- تابع پنهان سازی معادل دودویی و متنی که قرار است داده در آن مخفی شود را دریافت می کند.

۲- دو رقم دو رقم از رمز دودویی جدا می کند.

۳- تشخیص می دهد که این دو رقم مربوط به کدام مجموعه ذکر شده در جدول ۱-۳ است.

۴- در روی متن به دنبال حروفی می گردد که در مجموعه مشخص شده باشد.

۵- اگر حرف مورد نظر در متن پیدا شد:

اگر در گروه "۰۰" بود بدون بررسی هیچ شرطی دو کد "zwnj" بعد از آن قرار می دهیم.

اگر در گروه های دیگر بود و کاراکتر بعد از آن کاراکتر کشش با کد ۰۶۴۰ نباشد بررسی می شود که کاراکتر بعد از حرف جز حروف الفبا است یا خیر. اگر جز حروف الفبا بود از دو کاراکتر "zwj" و در غیر اینصورت از دو "zwnj" استفاده می شود.

۶- بعد از نشانه گذاری شمارنده روی متن و همچنین روی رمز دودویی جلو می رود.

۷- مراحل ۲ تا ۶ آنقدر تکرار می شوند تا کل رمز دودویی مخفی شود یا متن به پایان برسد.

۸- چون در هر بار عمل پنهان سازی دو رقم مخفی می شود در پایان نصف طول رمز دودویی که الگوریتم موفق به پنهان سازی شده به عنوان کلید برگردانده می شود. البته اگر طول رمز دودویی مخفی شده مضرب ۶ نباشد، بیت های اضافی نادیده گرفته شده تا طول آن مضربی از ۶ باشد به دلیل اینکه هر حرف معادل ۶ رقم دودویی است.

۳-۲-۵- روش بازیابی پیام

این روش عکس عمل پنهان سازی است. الگوریتم آن را به شرح زیر توضیح می دهیم:

۱- ابتدا کلیدی که از تابع پنهان سازی بدست آمده و متن استگو، به برنامه داده می شود.

۲- یک شمارنده با مقدار ۰ اختصاص داده می شود.

۳- برنامه متن استگو را مرور می کند تا به نشانه (zwnj+zwnj) یا (zwj+zwj) برسد. اگر این نشانه در متن پیدا شد، نشان دهنده اینست که در حرف قبل از این نشانه ها رمزی پنهان شده است.

۴- حرف مورد نظر استخراج شده و با گروه حروف موجود در جدول ۱-۳ مقایسه می شود.

۵- با تعیین گروهی که حرف متعلق به آنست مشخص می شود که چه داده ای در آن مخفی شده و دو رقم از داده بدست می آید.

۶- شمارنده یک واحد جلو می‌رود.

۷- مراحل ۳ تا ۶ آنقدر تکرار می‌شود تا مقدار شمارنده به مقدار کلید برسد. به این ترتیب کل رمز دودویی بدست می‌آید.

بعد از به دست آمدن رمز دودویی باتوجه به اینکه برای هر حرف ۶ رقم دودویی در نظر گرفته شده است. ۶ رقم ۶ رقم از پیام دودویی جدا شده، به مبنای ۱۰ برده می‌شود و مشخص می‌کند که عدد بدست آمده در مبنای ۱۰، شماره ترتیب کدامیک از حروف الفباست که در آرایه حروف به آن نسبت داده شده است. این کار تکرار می‌شود تا معادل کاراکتری پیام بدست آید و به این ترتیب رمز بدست می‌آید.

۳-۳- بررسی و نتیجه گیری

۳-۳-۱- بررسی ظرفیت

روش ارائه شده بر روی چندین فایل Word با فرمت doc و docx آزمایش شد. در این آزمایشات، طول متن اصلی و طول رمز متفاوت در نظر گرفته شد. جدول ۲-۱۱ این نتایج را نشان می‌دهد.

جدول ۲-۱۱: مقایسه ظرفیت روش در متون مختلف

طول متن (تعداد کاراکتر)	طول رمز (تعداد کاراکتر)	درصد ظرفیت (طول رمز / طول متن) * ۱۰۰
۵۷۱۲	۳۳۶	۵.۸۸
۵۲۴۲	۳۱۸	۶.۰۶
۵۱۱	۲۸	۵.۴۷
۷۸۴	۳۲	۴.۰۸
۳۸۲۱	۲۱۴	۵.۶۰
۱۰۷۰	۵۹	۵.۵۱
۱۵۰۹	۷۹	۵.۲۳
۷۶۹	۳۵	۴.۵۵
۱۶۴۶	۷۱	۴.۳۱
۲۱۳۶	۱۲۶	۵.۸۹
میانگین ظرفیت		۵.۲۵

همانطور که از داده های جدول ۲-۱۱ بدست می‌آید، متوسط ظرفیت این روش در حدود ۵.۲۵ درصد است که در بین روش های پنهان سازی درصد قابل قبولی است

مشاور: مؤسسه آموزش عالی روزبهان ساری	مرور و ارائه الگوریتم	کارفرما: وزارت ارتباطات و فناوری اطلاعات
تاریخ آخرین تجدید نظر: ۹۲/۸/۱۴	استکانوگرافی	نسخه: ۱

۳-۳-۲- آسیب پذیری در برابر حملات

بررسی های دیگر انجام گرفته نشان دهنده موارد زیر است:

۱- روش ما به دلیل استفاده از یونیکد، نسبت به تغییرات نوع، سایز و حالت فونت مقاوم است . حتی اگر قسمتی از فونت متن استگو را به یک نوع و قسمت دیگر را به نوع دیگری تبدیل کنیم داده از بین نمی رود.

۲- این روش برخلاف روش های استفاده از فضاهاى خالی، نسبت به اضافه یا کم کردن فضاهاى خالی مقاوم است. یعنی اگر فاصله بین خطوط را کم یا زیاد کنیم تا تورفتگی ای در پاراگراف ها ایجاد کنیم، داده از بین نمی رود.

۳- نسبت به تغییر قالب بندی متن مقاوم است. به عنوان مثال تغییر از حالت چپ چین به راست چین یا برعکس.

۴- اضافه کردن کلمه به متن در بین کلمات دیگر داده را از بین نمی برد. اما نباید در بین حروف یک کلمه چیزی اضافه نمود.

چون متن استگو ما درون فایل Word ذخیره می شود. اضافه کردن موارد دیگری مانند عکس نیز به متن استگو آسیبی وارد نمی کند.

۳-۳-۳- محدودیت ها و نواقص

این روش دارای محدودیت ها و نواقصی نیز می باشد که به آنها اشاره می کنیم:

۱- به دلیل استفاده از یونیکد، تنها برای متون الکترونیکی مناسب است و در متون چاپی قابل استفاده نیست.

۲- تنها کاراکترهای فارسی و فاصله را می توان در پیام رمز استفاده کرد و قابلیت پنهان سازی فایل هایی با نوع مختلف را ندارد. می توان با گسترش عملکرد آن این امکان را بوجود آورد .

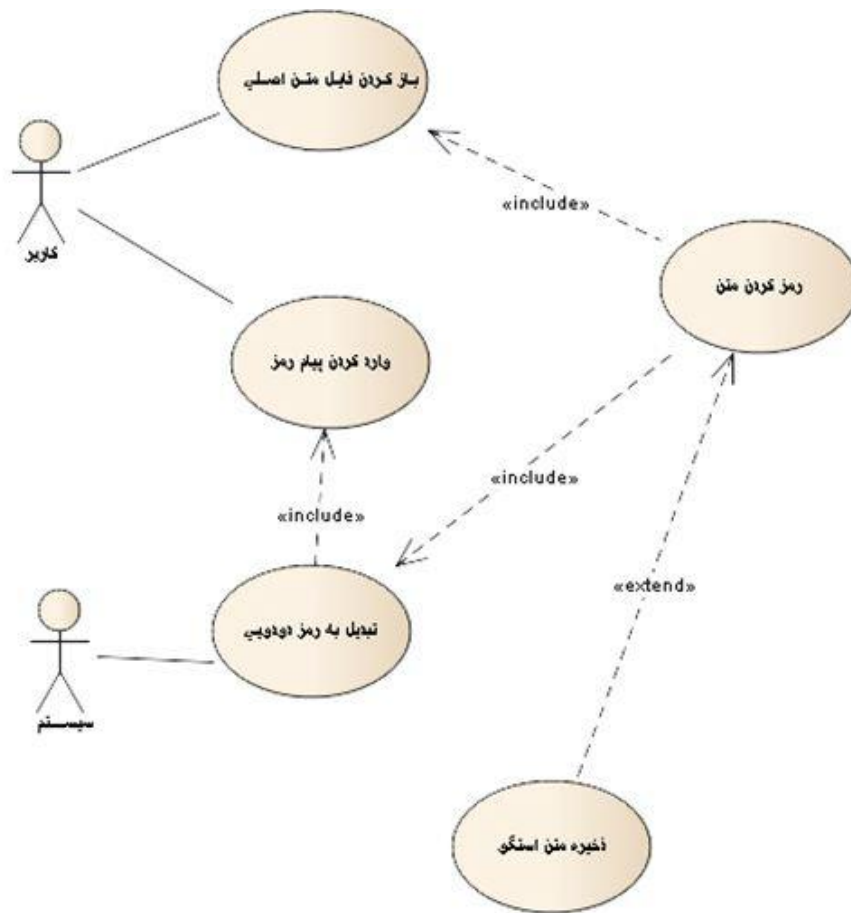
۳- حذف قسمتی از متن استگو، داده را از بین می برد.

۴- به دلیل اینکه در هر بار پنهان سازی دو رقم دودویی، دو کاراکتر به متن اضافه می شود. این کار می تواند موجب افزایش سایز فایل استگو شود.

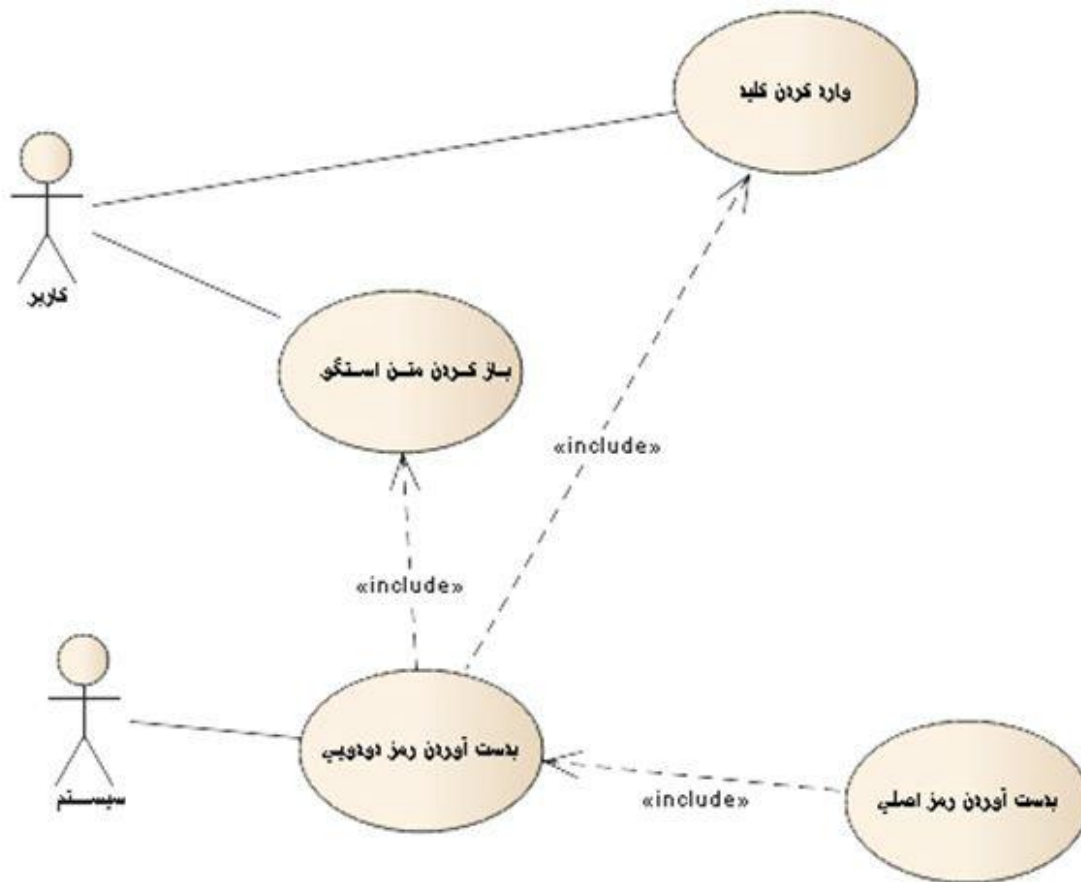
۳-۳-۴- ارزیابی کلی

همانطور که در فصول قبل نیز اشاره شد . روش های پنهان سازی باید سه ویژگی شفافیت، ظرفیت و مقاومت را به خوبی برآورده کنند که این موارد در روش ارائه شده بررسی شده است:

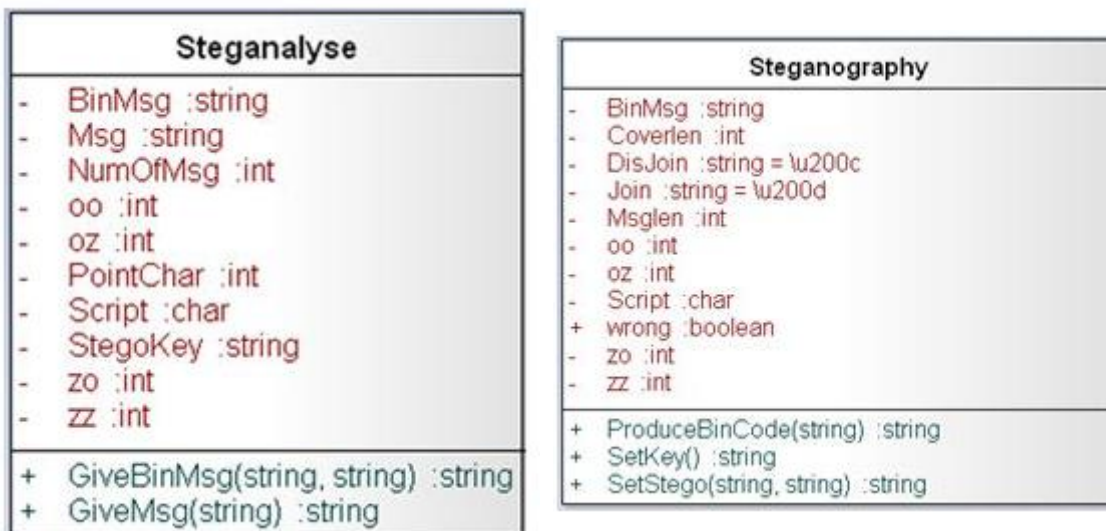
مشاور: مؤسسه آموزش عالی روزبهان ساری	مرور و ارائه الگوریتم	کارفرما: وزارت ارتباطات و فناوری اطلاعات
تاریخ آخرین تجدید نظر: ۹۲/۸/۱۴	استکانوگرافی	نسخه: ۱
<p>۱- شفافیت: به دلیل استفاده از کاراکترهای خاص در یونیکد، تغییر محسوسی در ظاهر متن اصلی ایجاد نمی‌کند.</p> <p>۲- ظرفیت: با توجه به بررسی های انجام شده، ظرفیت این روش در حدود ۵.۲۵ درصد است که رقم مناسبی می باشد.</p> <p>۳- استقامت: آزمایشات انجام گرفته و مواردی که در قسمت ۲-۳-۲ آورده شد. نشان دهنده اینست که این روش در مقابل خیلی از تغییرات مقاوم است.</p> <p>بنابراین روش ما ویژگی های یک روش پنهان سازی مناسب را دارا می باشد</p> <p>این روش به دلیل استفاده از یونیکد، قابل پیاده سازی در تمام فایل هایی است که از این سیستم پشتیبانی می کنند مانند فایل های Html.</p> <p style="text-align: center;">۳-۴- بررسی الگوریتم روش</p> <p>برای پیاده سازی الگوریتم این روش، از زبان برنامه نویسی C# ۲۰۰۸ استفاده شده است. برای پیاده سازی دو کلاس نوشته شد که یکی برای عمل پنهان سازی و دیگری برای عمل بازیابی پیام است. کد این دو کلاس در پیوست "الف" آورده شده است. همچنین در شکل ۲-۳ و ۲-۴ نمودار مورد کاربرد و در شکل ۲-۵ نمودار کلاس نشان داده شده است.</p>		



شکل ۲-۳: نمودار مورد کاربرد عملیات پنهان سازی



شکل ۲-۴: نمودار مورد کاربرد عملیات بازیابی پیام



شکل ۲-۵: نمودار کلاس