
بسم الله الرحمن الرحيم

دانشگاه صنعتی شاهرود

دانشکده آموزش های الکترونیکی

گروه مهندسی کامپیوتر و فناوری اطلاعات

عنوان:

نهان نگاری در متون فارسی با استفاده از روش های آماری

پژوهشگر:

هاجر ندیمی

استاد راهنما:

دکتر مرتضی زاهدی

استاد مشاور:

دکتر علی اکبر پویان

پایان نامه کارشناسی ارشد رشته مهندسی کامپیوتر گرایش هوش مصنوعی.

بهمن ماه ۱۳۹۴

دانشگاه صنعتی شاهرود

دانشکده : آموزش های الکترونیکی

گروه : مهندسی کامپیوتر و فناوری اطلاعات

پایان نامه کارشناسی ارشد خانم هاجر ندیمی

تحت عنوان: نهان نگاری در متون فارسی با استفاده از روش های آماری

در تاریخ ..... توسط کمیته تخصصی زیر جهت اخذ مدرک کارشناسی ارشد مورد ارزیـابی
و با درجه ..... مورد پذیرش قرار گرفت.

اساتید راهنما	امضاء	اساتید مشاور	امضاء
نام و نام خانوادگی :دکتر مرتضی زاهدی		نام و نام خانوادگی :دکتر علی اکبر پویان	
نام و نام خانوادگی :		نام و نام خانوادگی :	

اساتید داور	امضاء	نماینده تحصیلات تکمیلی	امضاء
نام و نام خانوادگی :		نام و نام خانوادگی :	
نام و نام خانوادگی :			
نام و نام خانوادگی :			
نام و نام خانوادگی :			

تقدیم به

روح پر فروغ مادرم که همواره یادش در دلم نور امید به زندگی را می تاباند.

پدر و همسر مهربانم که همیشه پشتیبان من بوده اند.

همه عزیزانی که دعای خیرشان بدرقه راهم بوده و هست.

تشکر و قدردانی

در ابتدا از استاد شایسته؛ جناب آقای دکتر زاهدی که در کمال سعه صدر، از هیچ کمکی در این عرصه بر من دریغ ننمودند و زحمت راهنمایی این رساله را بر عهده گرفتند؛ کمال تشکر و قدردانی را دارم .

همچنین از استاد ارجمند جناب آقای دکتر پویان ، که در انجام این رساله از راهنمایی های سودمند ایشان بهره مند شده ام ، کمال قدردانی را دارم .

و در پایان سپاس بی پایان خود را نثار همسر و پدر بزرگوارم که در تمام طول تحصیلات، همواره پشتیبان من بوده اند، مینمایم.

اینجانب هاجر ندیمی دانشجوی دوره کارشناسی ارشد رشته هوش مصنوعی دانشکده آموزش های الکترونیکی دانشگاه صنعتی شاهرود نویسنده پایان نامه نهان نگاری در متون فارسی با استفاده از روش های آماری تحت راهنمائی دکتر مرتضی زاهدی متعهد می شوم .

- تحقیقات در این پایان نامه توسط اینجانب انجام شده است و از صحت و اصالت برخوردار است .
- در استفاده از نتایج پژوهشهای محققان دیگر به مرجع مورد استفاده استناد شده است .
- مطالب مندرج در پایان نامه تاکنون توسط خود یا فرد دیگری برای دریافت هیچ نوع مدرک یا امتیازی در هیچ جا ارائه نشده است .
- کلیه حقوق معنوی این اثر متعلق به دانشگاه صنعتی شاهرود می باشد و مقالات مستخرج با نام « دانشگاه صنعتی شاهرود » و یا « Shahrood University of Technology » به چاپ خواهد رسید .
- حقوق معنوی تمام افرادی که در به دست آمدن نتایج اصلی پایان نامه تأثیرگذار بوده اند در مقالات مستخرج از پایان نامه رعایت می گردد.
- در کلیه مراحل انجام این پایان نامه ، در مواردی که از موجود زنده (یا بافتهای آنها) استفاده شده است ضوابط و اصول اخلاقی رعایت شده است .
- در کلیه مراحل انجام این پایان نامه، در مواردی که به حوزه اطلاعات شخصی افراد دسترسی یافته یا استفاده شده است اصل رازداری ، ضوابط و اصول اخلاق انسانی رعایت شده است .

تاریخ

امضای دانشجو

چکیده :

با توجه به گسترش استفاده از اینترنت ، تامین امنیت اطلاعات مبادله شده از اهمیت بسیار بالایی برخوردار است . جهت تحقق این امر یک روش این است که داده‌ها برای افراد غیرمجاز قابل‌فهم نباشدکه به آن رمزنگاری می‌گویند . روش دیگر مخفی نمودن ارتباط است که نهان‌نگاری نامیده می‌شود . در این رساله برای نهان‌نگاری از کاراکتر نیم‌فاصله به همراه کاراکتر فاصله استفاده‌شده است. برای فاصله بین حروفی که به حرف قبلی یا بعدی نمی‌چسبند نیز از نیم‌فاصله استفاده می‌شود تا ظرفیت نهان‌نگاری به طور میانگین ۱.۳۱ برابر روشی که تنها از کاراکتر نیم‌فاصله استفاده می‌شود ، افزایش یابد. همچنین با تغییر الگوی نهان‌نگاری ، مفاهیم بیت‌های صفر و یک را در طول متن تغییر می‌دهیم و از این طریق امنیت روش نهان‌نگاری را افزایش می دهیم . برای افزایش مقاومت نهان‌نگاری ابتدا پیام را چند بار در متن تکرار می‌کنیم، بدین‌صورت که درست بعد از اتمام پیام دوباره شروع به نهان نمودن پیام در ادامه متن برای چند بار می‌شود. تا در مرحله اکتشاف چندین بار پیام استخراج شود و نتایج باهم مقایسه شوند تا در صورت وجود خطا، به پیام درست دست‌یافته شود که این باعث افزایش مقاومت نهان‌نگاری می‌شود. علاوه بر این با ترکیب ویژگی‌های مختلف کاراکتر نیم‌فاصله (نوع، سایز، رنگ، ضخامت و خمیدگی قلم)، مقاومت نهان‌نگاری افزایش داده می‌شود. در این حالت نتایج اکتشاف برای تمام ویژگی‌ها کنار هم گذاشته می‌شود تا بتوان به پیام درست دست‌یافت. لذا با این دو روش مقاومت نهان‌نگاری افزایش‌یافته طوری که می‌توان به مقاومت ۹۹.۷۷٪ رسید که نشان‌دهنده مقاومت بسیار بالای روش است.

کلمات کلیدی:نهان نگاری،کاراکتر نیم فاصله،مقاومت نهان نگاری،ظرفیت نهان نگاری،امنیت نهان نگاری

فهرست مطالب

فصل اول : مقدمه

فصل دوم : نهان‌نگاری

فصل سوم : روش پیشنهادی

فصل چهارم : نتایج نهان‌نگاری

فصل پنجم : نتیجه‌گیری و پیشنهادات

فهرست اشکال

فهرست جداول

فصل اول : مقدمه

۱-۱ - مقدمه :

با توجه به گسترش استفاده از اینترنت ، نیاز به یک بستر ارتباطی امن، جهت تبادل داده بیش‌ازپیش احساس می‌شود. برای برقراری این ارتباط امن می‌توان داده‌ها را به طریقی انتقال داد که برای افراد غیرمجاز قابل‌فهم نباشد ، که رمزنگاری[۱] نامیده می‌شود و روش دیگر مخفی نمودن ارتباط است که به نهان‌نگاری[۲] معروف است [۱] .

از نهان‌نگاری، برای بالا بردن حفاظت و امنیت اطلاعات استفاده می‌شود. و هدف آن ، پنهان کردن اطلاعات به‌گونه‌ای است که تنها افراد مجاز از وجود ارتباط و اطلاعات مطلع باشند .برخلاف رمزنگاری که اولویت آن مخفی نمودن محتوای پیام است در نهان‌نگاری هدف پنهان کردن خود ارتباط است . مفهوم از امنیت در رمزنگاری محرمانگی پیام است اما امنیت در نهان‌نگاری نامحسوس بودن حضور پیام در سیگنال میزبان می‌باشد . در نهان‌نگاری،به‌عنوان رسانه پوششی از داده‌هایی مانند تصویر، صوت، ویدئو و متن استفاده می‌شود.

بنا به نظریه شانون در زمینه سیستم‌های محرمانه ، سه دسته ارتباط وجود دارد [۲] :

۱. سیستم اختفا[۳](روش‌هایی که وجود پیام از دید دشمن مخفی است)
۲. سیستم پوشیدگی[۴]
۳. سیستم رمزنگاری

شانون نهان‌نگاری را از زیرشاخه‌های سیستم‌های اختفا می‌داند .

در مقابل نهان‌نگاری، مبحث نهان کاوی[۵] (تحلیل نهان‌نگاری) مطرح می‌شود که عمده‌ترین هدف آن تشخیص وجود اطلاعات پنهان در رسانه مشکوک می‌باشد . علاوه بر این ، استخراج ،

[1] Cryptography
[2] Steganography
[3] Concealment System
[4] Privacy Systems
[5] Steganalysis

تغییر و حذف اطلاعات نهان نیز از اهداف نهان کاوی است که البته کمتر موردتوجه تحلیلگران بوده است؛ زیرا تنها با تشخیص وجود اطلاعات ، یک روش نهان‌نگاری، شکست می‌خورد.

در این پژوهش به بررسی دقیق انواع روش‌های نهان‌نگاری و ارزیابی آن‌ها و بررسی نواقص هرکدام از روش‌ها ازجمله ظرفیت پایین و قابل‌شناسایی بودن توسط روش‌های نهان کاوی ، می‌پردازیم .

۱-۲ - تاریخچه

نهان‌نگاری معادل واژه Steganography است که در اصل واژهای یونانی بوده و از ترکیب دو کلمه Steganos به معنای پوشش و Grpahy به معنی نوشتن تشکیل‌شده است [۳] .

نهان‌نگاری به‌عنوان یک هنر از قدیمی‌ترین فنونی است که انسان به آن مشغولیت یافته است. سابقه اولین طرح‌های پنهان کردن اطلاعات از دید دشمن، به حدود ۴۰۰۰ سال پیش بازمی‌گردد [۴] . اما نهان‌نگاری به‌عنوان یک علم بسیار جوان است و از سه دهه گذشته موردتوجه محققین حوزه ارتباطات واقع‌شده است و تحقیقات زیادی در این زمینه صورت می‌گیرد.

قدیمی‌ترین مثال نهان‌نگاری به حدود سال‌های ۴۴۰ قبل از میلاد برمی‌گردد . هنگامی‌که حاکم یونان به دست داریوش زندانی‌شده بود، به دنبال راهی می‌گشت تا پیام‌های مخفی را به لشکریان خودی برساند .او سر برده‌ها را می‌تراشید ، پیام را روی سر آن‌ها خال‌کوبی می‌کرد و پس از رشد مجدد موها، برده‌ها را عازم مقصد می‌نمود.

نهان‌نگاری در قرن‌های ۱۵ و ۱۶ توسعه یافت .یکی از رساله‌هایی که در این زمینه نوشته‌شده توسط ویلکینز[۶] است، که بعداً در کالج ترینتی[۷] به استادی رسید .او روش‌هایی را از کد کردن

[6] Wilkins
[7] Trinity

پیام‌ها در موزیک تا جوهرهای نامرئی پیشنهاد داد. همچنین او اولین طرح‌ها را در رمزگشایی با استفاده از تناوب کلمات ساخت [۵] .

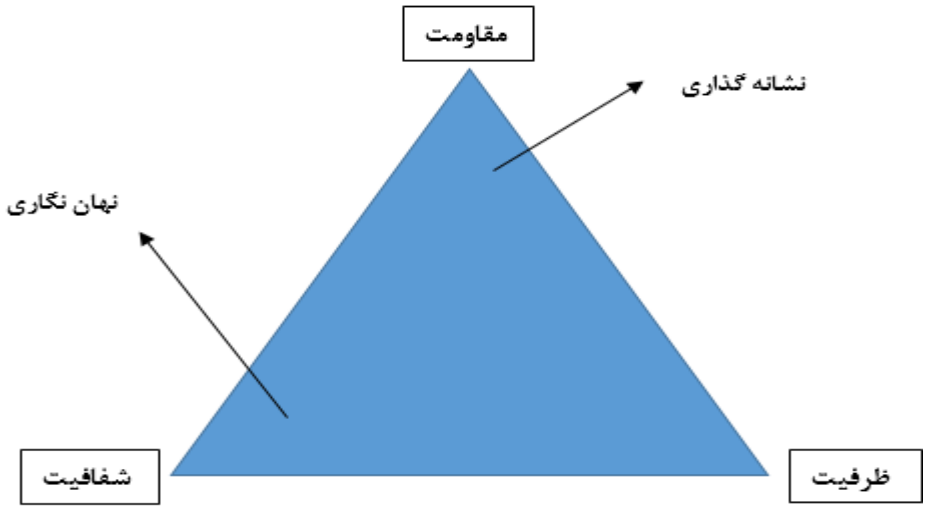
اولین روش‌های نهان‌نگاری دیجیتال در دهه ۸۰ مطرح شدند . در سال ۱۹۸۴ تعریف نهان‌نگاری به شکل کلاسیک توسط سیمونز[۸] تحت عنوان مساله زندانی‌ها بیان شد [۶] . آلیس و باب دو زندانی در دو سلول جداگانه هستند که قبل از زندانی شدن، توافقاتی با یکدیگر به‌منظور طراحی نقشه فرار انجام داده‌اند و بر روی یک کلید مشترک توافق کرده‌اند .آن‌ها می‌توانند پیغام‌هایی را با یکدیگر مبادله کنند .هر ارتباطی که بین این دو صورت می‌گیرد توسط وندی که زندانبان است کنترل می‌شود .وندی اگر فعالیت شک‌برانگیزی را مشاهده کند، مانع ارتباط باب و آلیس می‌شود . پس پیغام‌های مبادله شده باید طوری باشند که وندی را مشکوک نسازد .بنابراین باب و آلیس از نهان‌نگاری استفاده کرده و پیام‌های نقشه فرار خود را در اشیاء پوششی مناسبی مانند تصاویر پنهان می‌کنند . شیء حاوی پیغام که به آن نهان نگاره میگویند از طریق کانال ناامنی که وندی ناظر آن است و به آن دسترسی دارد، بین دو طرف مبادله می‌شود .وندی در این مسأله یک مهاجم است و می‌تواند فعال و یا غیرفعال باشد .اگر مهاجم فعال باشد، می‌تواند محتوای پیام مورد مبادله را تغییر دهد و نیز می‌تواند پیامی را جعل کرده و به‌جای آلیس به باب و یا از باب به آلیس بفرستد .در حالتی که مهاجم غیرفعال است، تنها ضرورت برای برقراری امنیت روش نهان‌نگاری این است که داده جایگذاری شده غیرقابل‌شناسایی باشد.

تقریباًهمه کارهای انجام‌شده در این زمینه روش‌های نهان‌نگاری دیجیتال مربوط به دو دهه اخیر است .

[8]Simmons

۳-۱ پارامترهای سیستم نهان‌نگاری اطلاعات

برای اندازه‌گیری میزان کارایی یک سیستم نهان‌نگاری اطلاعات سه پارامتر اساسی موردبررسی قرار می‌گیرند که عبارت‌اند از :شفافیت[۹] ، ظرفیت[۱۰] و مقاومت[۱۱] . این پارامترها را می‌توان در سه رأس یک مثلث قرارداد و به‌طور هم‌زمان دسترسی به این پارامترها امکان‌پذیر نیست و در تقابل با یکدیگرند . این الزامات به‌عنوان "مثلث سحر و جادو" شناخته می‌شود و متناقض هستند. سیستم‌های نهان‌نگاری بیشتر در نزدیکی رأس چپ و سیستم‌های نشانه‌گذاری در حوالی رأس بالا قرار می‌گیرند[۷] .



شکل (۱-۱) پارامترهای سیستم نهان‌نگاری اطلاعات[۷]

- شفافیت : شفافیت به معنای آن است که سیگنال نهان‌نگاری شده، در اثر نهان‌نگاری دچار تغییرات محسوس و آشکارنشده باشد و تا حد امکان به سیگنال پوشش اولیه شبیه باشد.

امنیت در یک سیستم نهان‌نگاری مترادف با میزان شفافیت آن است . در یک سیستم نهان‌نگاری امن ، کشف حضور پیام مخفی در رسانه پوشش توسط روش‌های مختلف امری مشکل می‌باشد .اگر

[9] Transparency
[10] Capacity
[11] Robustness

دقت شناسایی یک روش نهان‌نگاری کمتر از ۶۰ ٪ باشد ، می‌توان الگوریتم نهان‌نگاری را غیرقابل‌شناسایی و امن در نظر گرفت [۸] .

- ظرفیت : نهایت میزان اطلاعاتی است که در رسانه پوششی قابل جایگذاری می‌باشد به‌گونه‌ای که نتوان با روش‌های نهان کاوی آن را تشخیص داد .
- مقاومت : مقاومت الگوریتم‌های نهان‌نگاری ، نیاز اصلی یک سیستم نهان‌نگاری اطلاعات نیست، چون هدف این سیستم‌ها مخفی نگه‌داشتن حضور پیام پنهان است.

مقاومت در مقابل حملات ، در سیستم‌های نشانه‌گذاری دارای اهمیت و توجه است .در نشانه‌گذاری هدف رسیدن به سطح بالایی از مقاومت است، که باوجود دانستن وجود پیام ، نتوان آن را تغییر داد و یا تخریب کرد.

۱-۴ تعاریف و اصطلاحات

تعاریفی که در اولین کارگاه بین‌المللی نهان‌نگاری اطلاعات برای انجام این عمل عنوان شد به شرح زیر است: [۹]

۱) رسانه پوششی [۱۲] : شیء ای که پیام در قالب آن منتقل می‌شود و می‌تواند شامل تصویر ،متن،صوت و ویدئو و ... باشد که به آن رسانه میزبان هم گفته می‌شود.

۲) پیام جاسازی‌شده[۱۳] : داده‌ای که باید به‌صورت پنهانی منتقل شود و داخل رسانه میزبان جاسازی می‌گردد.

۳) رسانه استگو[۱۴] : حاصل ترکیب پیام در میزبان است.

۴) کلید استگو[۱۵] : اطلاعات سری که مشترک بین فرستنده و گیرنده است و به‌منظور جاسازی و بازیابی اطلاعات از آن استفاده می‌شود.

۵) جاسازی کننده[۱۶] : تابع جاسازی کننده پیام.

۶) استخراج‌کننده[۱۷] : تابع بازسازی‌کننده پیام .

[12] cover-medium
[13] Embedded-message
[14] Stego-medium
[15] Stego-key
[16] Embedder(E)
[17] extractor(E-1)

۱-۵ چشم‌انداز

در این پایان‌نامه ابتدا روش‌های مختلف نهان‌نگاری در متون موردبررسی قرار می‌گیرد و کارایی الگوریتم‌ها ازلحاظ مقاومت ، ظرفیت ، امنیت و تناسب با رسانه حامل مورد ارزیابی قرار می‌گیرد . و درنهایت سعی خواهد شد با ترکیب روش‌های موجود و بهره‌گیری از روش‌های آماری الگو،الگوریتمی ارائه شود که پارامترهای لازم را در حد امکان به مقدار مطلوب برساند.

در ادامه این نوشته ابتدا در فصل دوم به‌تفصیل در مورد نهان‌نگاری و تعاریفی که در این زمینه وجود دارد و انواع الگوریتم‌های نهان‌نگاری و نقاط ضعف و قوت هریک، بحث خواهد شد . در فصل سوم به توضیح روش‌های نهان‌نگاری پیشنهادی خواهیم پرداخت و در فصل چهارم کارایی روش‌های پیشنهادی مورد ارزیابی واقع می‌شود و نتایج آزمایشات انجام‌شده مورد تجزیه‌وتحلیل قرار می‌گیرد. جمع بندی نهایی و پیشنهادهایی برای کارهای آینده در فصل پنجم ارائه خواهد شد.

فصل دوم : نهان‌نگاری

۲-۱ نهان‌نگاری چیست ؟

نهان‌نگاری که نام لاتین آن استگانوگرافی می‌باشد به معنی پنهان کردن وجود ارتباط می‌باشد به این صورت که پیام مخفی را در یک رسانه پوششی قرار می دهند طوری که تخریب و تغییر بوجودآمده مشهود نبوده و افراد غیرمجاز متوجه وجود ارتباط نشوند .

نهان‌نگاری یکی از شاخه‌های ارتباطات پوشیده می‌باشد ، از دیگر شاخه‌های آن رمزنگاری می‌باشد . در رمزنگاری غیرقابل دسترسی بودن محتوای پیام برای ما حائز اهمیت می‌باشد و هدف حفظ محرمانگی اطلاعات است اما در نهان‌نگاری وجود ارتباط انکار می‌شود .

قبل از پنهان نمودن پیام با استفاده از الگوریتم‌های رمزنگاری آن را رمز کرده و سپس نهان‌نگاری انجام می‌شود . درنتیجه زمان کشف و استخراج پیام برای دسترسی به آن باید از سه لایه حفاظتی گذشت . ابتدا باید وجود این ارتباط نامحسوس کشف شود که با توجه به اینکه تغییراتی که به وجود می‌آید بسیار ناچیز است و اصولاً تغییر به وجود آمده مشهود نیست نیاز به زمان دارد ، بعد از کشف آن باید الگوریتم نهان‌نگاری را به دست آورد که با توجه به الگوریتم‌های پیچیده و کلیدهایی که تعریف می‌شود فرایند دشواری خواهد بود . اگر از دو مرحله قبل با موفقیت بگذرد و متن رمزنگاری‌شده به دست گیرنده غیرمجاز برسد مسئله رمزنگاری مطرح می‌شود .

۲-۲ نهان کاوی چیست ؟

در کنار گسترش و پیشرفت الگوریتم‌های نهان‌نگاری ، روش‌های نهان کاوی (تحلیل نهان‌نگاری) که وجود پیام مخفی را در رسانه پوششی کشف می‌کند ارتقاءیافته‌اند . نهان کاوی تشخیص درست وجود پیام مخفی توسط الگوریتم‌های مختلف در رسانه میزبان است . درواقع، الگوریتم‌های تحلیل نهان‌نگاری امنیت الگوریتم‌های نهان‌نگاری را بررسی می‌کنند .

الگوریتم‌های تحلیل نهان‌نگاری از روش‌های تحلیل آماری و یا ماشین‌های یادگیرنده [۱۸] ، برای شناسایی حضور و یا عدم حضور پیام مخفی در رسانه پوششی استفاده می‌کنند .یک الگوریتم نهان‌نگاری ناموفق است اگر الگوریتم‌های نهان کاوی بتوانند با تخمین % 50 ، پی به وجود پیام پنهان ببرند حتی اگر محتوای پیام را نتوانند کشف کنند.

با توجه به اینکه چه میزان اطلاعات راجع به پیام ادغام‌شده در اختیار است، میزان اثربخشی تحلیل نهان‌نگاری متفاوت است. روش‌های حمله نهان‌نگاری به شش نوع تقسیم می‌شوند [۱۰] :

- حمله فقط نهان‌نگاری [۱۹]: تنها فایلی که داده‌ی ادغام‌شده دارد در دسترس است.
- حمله حامل شناخته‌شده[۲۰] : هم فایل حامل اصلی و هم فایل نهایی (که پیام در آن ادغام‌شده) در دسترس هستند.
- حمله پیام شناخته‌شده [۲۱] : پیام اصلی قبل از ادغام شدن در حامل، موجود است.
- حمله نهان‌نگاری انتخاب‌شده[۲۲] : هم الگوریتم استفاده‌شده برای ادغام کردن داده و هم فایل نهایی موجود هستند.

[18] Learning Machine
[19] Chosen-steganography attack
[20] Chosen-message attack
[21] Known-message attack
[22] Chosen-steganography attack

- حمله پیام انتخاب‌شده [۲۳]: پیام اصلی و الگوریتم استفاده‌شده برای ادغام پیام موجود هستند اما فایل حامل و فایل نهایی در دسترس نیستند. این حمله توسط تحلیل‌گر برای مقایسه با فایل‌های آینده استفاده می‌شود.
- حمله نهان‌نگاری شناخته‌شده[۲۴] : همه‌ی عناصر سیستم (پیام اصلی، حامل پیام و الگوریتم) برای تحلیل در دسترس هستند.

طبق اصل Kerckhoffs که اصل پذیرفته‌شده‌ای در رمزنگاری است ، امنیت سیستم رمزنگاری تنها به مخفی بودن کلید و نه کمبود اطلاعات از الگوریتم بستگی دارد .اصل Kerckhoffs قابل‌تعمیم به نهان‌نگاری نیز هست و نپذیرفتن این اصل در نهان‌نگاری و مخفی نگه‌داشتن الگوریتم نهان‌نگاری و متکی کردن امنیت سیستم به مخفی بودن الگوریتم باعث پایین آمدن سطح امنیت سیستم می‌شود. روشن است که مشخص بودن الگوریتم برای تحلیل، احتمال موفقیت تحلیلگر را بالا خواهد برد.

اما عده‌ای نیز براین باورند که متکی کردن امنیت سیستم نهان‌نگاری به مخفی بودن فقط کلید، سبب پایین آمدن امنیت خواهد شد .بنابراین مخفی بودن الگوریتم تنها راه چاره برای امن نگه‌داشتن سیستم است و درنتیجه نباید امنیت نهان‌نگاری را تنها به مخفی بودن کلید وابسته کرد[۱۱] .

[23] Chosen-message attack
[24] Known-steganography attack

۲-۳ نهان‌نگاری در متن

برخلاف رسانه‌های دیگر چون تصویر ، صدا و کلیپ‌های ویدئویی ، استفاده از سندهای متنی از گذشته‌های دور رواج داشته و پس از اختراع ماشین چاپ نیز همچنان بسیاری از کتب و سندها تنها حاوی متن بوده‌اند .

امروزه با پیشرفت تکنولوژی ، حجم اطلاعات سمعی و بصری نیز افزایش چشمگیری داشته است ، ولی هنوز هم این حقیقت پابرجاست که کماکان اسناد متنی حجم عظیمی از اطلاعات را به خود اختصاص داده‌اند . به همین علت نهان‌نگاری در متن توجه محققین قرارگرفته است . نهان‌نگاری درواقع به مفهوم جاسازی کردن اطلاعات مهم در یک رسانه حامل است ، به‌نحوی‌که قابل‌تشخیص برای دیگران نباشد [۱۲] .

۲-۳-۱ اهداف نهان‌نگاری در متن

اهداف نهان‌نگاری در متن را می‌توان به دودسته زیر تقسیم کرد [۱۳] :

۱- پنهان‌سازی داده‌ها به‌منظور رساندن یک پیام محرمانه به یک شخص خاص

۲- درج امضاء خالق یک اثر در داخل آن ، به‌منظور حفظ حق نشر[۲۵]

[25] copyright

۲-۳-۲ انواع نهان‌نگاری

۲-۳-۲-۱ شیفت[۲۶] دادن خطوط

شیفت دادن خطوط به دو صورت افقی و عمودی انجام می‌گیرد . درروش افقی شیفت به سمت چپ یا راست صورت می‌پذیرد ، به این صورت که اگر خط به سمت راست شیفت پیدا کند بیت ۱ جهت نهان‌نگاری استفاده می‌شود و اگر جابجایی وجود نداشت بیت صفر[۱۴] . در شکل(۲-۱)نمونه‌ای از این روش نشان داده‌شده است . تصویر سمت راست ، متن حامل قبل از نهان‌نگاری و تصویر سمت چپ نتیجه نهان‌نگاری بیت‌های ۰۱۰۰۱۰۱۰ در متن حامل است .

با گسترش اینترنت و رشد سریع و روزافزون دادههای دیجیتال، نیاز به یک بستر ارتباطی امن، که در آن داده ها به طریقی امن مبادله شوند، بیش از پیش احساس میشود .در جهت برقراری یک ارتباط امن میتوان داده ها را به طریقی انتقال داد که برای شنونده غیر مجاز قابل فهم نباشد .طریق دیگر مخفی نمودن ارتباط است .روش اول به رمزنگاری موسوم است و روش دوم نهان نگاری نامیده می شود.

با گسترش اینترنت و رشد سریع و روزافزون دادههای دیجیتال، نیاز به یک بستر ارتباطی امن، که در آن داده ها به طریقی امن مبادله شوند، بیش از پیش احساس میشود .در جهت برقراری یک ارتباط امن میتوان داده ها را به طریقی انتقال داد که برای شنونده غیر مجاز قابل فهم نباشد .طریق دیگر مخفی نمودن ارتباط است .روش اول به رمزنگاری موسوم است و روش دوم نهان نگاری نامیده می شود.

شکل(۲-۱) نمونه‌ای از نهان‌نگاری به روش شیفت خطوط

درروش عمودی خطوط متن ، بر اساس مقدار بیت واترمارک (بیت ورودی) به سمت بالا یا پایین شیفت داده می‌شوند ، بدین ترتیب که از هر سه سطر متوالی ، سطر میانی (خط زوج)، به‌عنوان خط نامزد برای شیفت و دو سطر مجاور آن (خطوط فرد) به‌عنوان خطوط کنترلی در نظر گرفته و شیفت پیدا نمی‌کنند . خطوط کنترلی جهت عملیات آشکارسازی مورداستفاده قرار می‌گیرند تا بتوان با داشتن موقعیت آن‌ها ، موقعیت خط میانی را به‌صورت نسبی به دست آورد[۱۵] .

[26] shift

برای آشکارسازی تکنیک کدینگ شیفت خط ، می‌توان از روش‌های زیر استفاده نمود .

۲-۳-۲-۱-۱ روش پایه خطوط

در این روش فاصله پایه خط وسطی از پایه دو خط کنترلی مجاور محاسبه‌شده و پس از مقایسه با یکدیگر ، مقدار و جهت شیفت به دست می‌آید[۱۶] .

$$\text{If } l_2 - l_1 < l_3 - l_2 \longrightarrow \text{line 2 shifted up} \qquad (۱)$$

$$\text{If } l_2 - l_1 > l_3 - l_2 \longrightarrow \text{line 2 shifted down}$$

این روش فقط هنگامی کاربرد دارد که فاصله بین خطوط متن اصلی برابر باشند و برای متونی که دارای خطوط بافاصله نامنظم هستند کاربرد ندارد .

۲-۳-۲-۱-۲ روش مرکز ثقل[۲۷] خطوط

در این تکنیک ، ابتدا مرکز ثقل خطوط در سند اصلی با استفاده از رابطه (۲)محاسبه می‌شود [۱۶].

$$c_i = \left(\sum_{y=b_i}^{e_i} yh(y)\right) \Big/ \left(\sum_{y=b_i}^{e_i} h(y)\right) \qquad (۲)$$

برای آشکارسازی ، فاصله مرکز ثقل خطوط حامل از مراکز ثقل خطوط کنترلی مجاور در سند واترمارک شده محاسبه و پس از مقایسه این فواصل با فواصل متناظر در سند اصلی ، اطلاعات مخفی‌شده آشکار می‌شوند .

مزایا : به ازای مقدار شیفت یکسان در خطوط متون فارسی و لاتین ، اثر عملیات واترمارکینگ روی متون فارسی نامحسوس‌تر است ، همچنین قابلیت استفاده برای تمام زبان‌ها وجود دارد .

[27] Center of gravity

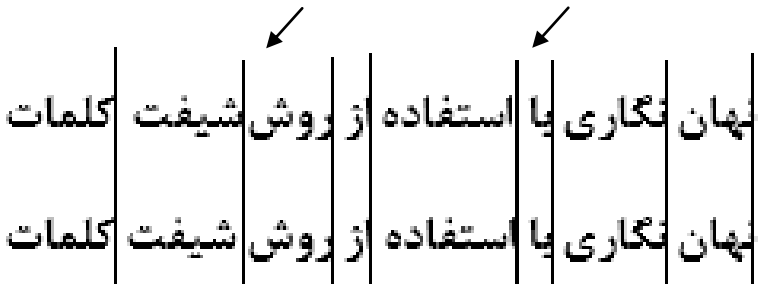
معایب : احتمال خطای آشکارسازی برای متون لاتین کمتر است .

۲-۳-۲-۲ شیفت دادن کلمات

در این روش کلمات حامل (زوج) به‌اندازه یک پیکسل به چپ یا راست شیفت داده می‌شوند . اگرچه تعداد بیت‌های واترمارک شده درروش شیفت کلمات بسیار بیشتر از شیفت خط می‌باشد اما در عوض تعداد خطای آشکارسازی نیز افزایش می‌یابد . با این روش طول خط حفظ می‌شود . در این روش به ازای تعداد کلمات بیت جهت نهان‌نگاری وجود دارد[۱۷].

با دو روش قابل‌دیدن و شناسایی است . اگر شخصی از الگوریتم فاصله‌ها اطلاع داشته باشد می‌تواند متن فعلی را با الگوریتم مقایسه کند و با استفاده از اختلاف‌ها اطلاعات مخفی‌شده را پیدا کند . روش دوم بررسی نقطه‌به‌نقطه عکسِ متن است تا فاصله‌های تغییریافته مشخص شود .

در شکل (۲-۲) نمونه‌ای از خروجی نهان‌نگاری با این روش نمایش داده‌شده است لازم به ذکر است این تصویر چند برابر بزرگ‌تر از تصویر اصلی نمایش داده‌شده است . خط دوم مربوط به نهان‌نگاری رشته بیت ۰۰۰۱۰۰۱۰ می‌باشد .



شکل(۲-۲) نمونه‌ای از نهان‌نگاری به روش شیفت کلمات

در این روش چنانچه شخص از الگوریتم فاصله‌ها اطلاع داشته باشد ، می‌تواند متن فعلی را با الگوریتم مقایسه کند و با استفاده از اختلاف ، اطلاعات پنهان‌شده را استخراج کند . همچنین می‌توان به بررسی دقیق تصویر متن پرداخت تا فواصل تغییریافته شناسایی شوند. این عمل گرچه بسیار

زمان‌بر می‌باشد ، اما امکان یافتن اطلاعات نهان شده در متن بسیار زیاد است . در این روش نیز با تایپ مجدد متن یا استفاده از برنامه‌های تشخیص حروف ، اطلاعات نهان شده از بین می‌رود .

مزایا : نسبت به روش شیفت خطوط ظرفیت بالاتری دارد ، تغییرات ناچیزی در متن ایجاد می‌شود و به زبان خاص وابسته نیست .

معایب : امکان شناسایی تغییرات ایجادشده توسط برخی OCR[۲۸] ها و اسکنرها وجود دارد و در مقابل قالب‌بندی مجدد صفحه مقاومت کمی دارد .

۲-۳-۲-۳ استفاده از ویژگی حروف

در این روش از خصوصیات حروف جهت نهان‌نگاری استفاده می‌کنیم . این خصوصیات می‌تواند شامل طول برخی کاراکترها ، نقاط کاراکترها و شیب بعضی از حروف باشد. به‌عنوان‌مثال درزمینه حروف نقطه‌دار ، ازآنجایی‌که تعداد حروف دارای نقطه در متون فارسی زیاد است ، این روش از نقاط حروف برای درج اطلاعات استفاده می‌کند . بدین‌صورت که برای درج بیت صفر مکان نقطه در حروف دارای نقطه تغییر نمی‌کند و برای درج بیت یک ، مکان نقطه کمی به سمت بالا انتقال می‌یابد. از ویژگی‌های این روش ظرفیت بالا و استحکام پایین می‌باشد[۱۸] .

شکل(۲-۳) نمونه‌ای از نهان‌نگاری به روش ویژگی حروف

[28] Optical Character Recognition

در این روش با قرار دادن حروف در یک‌فاصله ثابت با بقیه‌ی حروف ، اطلاعات از بین می‌روند . همچنین با تایپ مجدد متن یا استفاده از برنامه‌های تشخیص حروف اطلاعات نهان‌نگاری شده از بین می‌روند .

۲-۳-۲-۴ استفاده از حروف کشیده

این روش در زبان‌هایی استفاده می‌شود که کاراکترهای تشکیل‌دهنده یک کلمه به یکدیگر می‌چسبند . در این روش بیت صفر را حرف بدون نقطه که قبل از آن از کاراکتر کششی استفاده‌شده اختصاص می‌دهیم و بیت یک را به حرف نقطه‌دار که قبل از آن از کاراکتر کششی استفاده‌شده اختصاص می‌دهیم . شکل (۲-۴) خروجی این روش را برای نهان‌نگاری رشته بیت ۰۱۰۱۱۱۰ در متن حامل نشان می‌دهد

[۱۹] .

بیت محرمانه	**0101110**
متن پوششی	شکست در یک سیستم نهان‌نگاری به معنای پی بردن به وجود پیام مخفی می‌باشد
متن خروجی	شکـست در یک سـیسـتم نهان نگاری به معـنای پـی بردن به وجود پـیام مخفی مـی باشد . 0 1 0 1 1 1 0

شکل(۲-۴) نهان‌نگاری به روش استفاده از حروف کشیده

یکی از نگرانی‌هایی که در این روش وجود دارد این است که هنگامی اطلاعاتی که قرار است مخفی شود کمتر از متن پوششی باشد ، حروف در چند سطر ابتدایی دچار تغییر شده و در مابقی

بدون تغییر باقی می‌ماند ، این باعث آشکار شدن پیام مخفی می‌شود . این مشکل با ایجاد یک کاراکتر خاص به نام " کاراکتر پایانی " برطرف خواهد شد . این کاراکتر دارای کد ۱۱۱۱۱۱ بوده و بعد از درج آخرین بیت از پیام محرمانه تعبیه خواهد شد .

۲-۳-۲-۵ استفاده از فتحه در زبان عربی

این روش تنها در زبان عربی استفاده می‌شود . در این روش با تغییر دادن شکل فتحه عملیات نهان‌نگاری انجام می‌شود . به این صورت که اگر بخواهیم بیت ۱ را نهان‌نگاری کنیم فتحه را با شیب معکوس درج می‌کنیم و جهت نهان‌نگاری بیت ۰ فتحه را عادی درج می‌کنیم [۲۰].

این روش ظرفیت قابل قبولی دارد و از معایب آن می‌توان به ایجاد اثرات کاملاً محسوس در متن ، وابستگی کامل زبان و مقاومت بسیار پایین اشاره کرد .

۲-۳-۲-۶ استفاده از کلمات مترادف[۲۹]

در این روش تفاوت اندکی بین معنی متن اصلی و متن نهان‌نگاری شده وجود دارد . غنی بودن زبان فارسی ازلحاظ کلمات و عبارات مترادف باعث ایجاد دو مزیت عمده در استفاده از این زبان برای نهان‌نگاری متن گردیده است . نخست آنکه نهان‌نگاری متن با استفاده از کلمات و عبارات مترادف در این زبان تغییر چندانی در مفهوم متن حاصل ایجاد نمی‌نماید و دوم آنکه بالا بودن تعداد کلمات و عبارات معادل در این زبان ، حجم اطلاعات قابل نهان‌نگاری را در متن‌های زبان فارسی بالا می‌برد[۱۸]. کلمات و عبارات معادل در زبان فارسی را می‌توان در ۱۰ دسته به شرح زیر طبقه‌بندی نمود :

۱. کلمات با دو نگارش متداول
۲. کلمات بیگانه رایج

[29] Synonym

۳. کلمات دارای دو نوع جمع

۴. کلمات دارای دو فرم هم‌معنی

۵. کلمات مرکب مترادف

۶. کلمات کاملاً مترادف فارسی

۷. ترکیبات فارسی معادل کلمات عربی

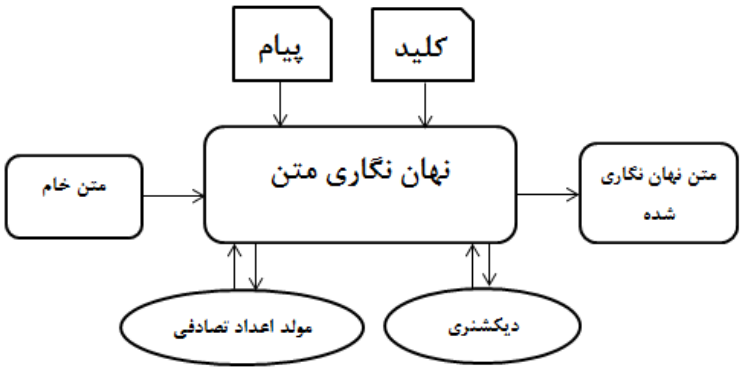
۸. عبارات معادل

۹. کلمات معادل در ترکیب‌های خاص

۱۰. فعل‌های معین هم‌معنی

فرآیند کلی نهان‌نگاری با استفاده از کلمات معادل و مترادف در شکل (۲-۵) نشان داده‌شده است

.



شکل(۲-۵)فرایند کلی نهان‌نگاری با استفاده از کلمات مترادف

مراحل استخراج اطلاعات نهان‌نگاری شده به شرح زیر هست .

- ارزیابی متن برای یافتن کلمات و عبارات کاملاً مترادف با استفاده از دیکشنری کلمات و عبارات مترادف

- تعیین کلمه موردنظر از میان لیست کلمات و عبارات مترادف موجود در متن ، برای استخراج بیت مربوطه با استفاده از کلید مخفی و مولد تصادفی که در هنگام نهان‌نگاری اطلاعات بکار گرفته‌شده‌اند .
- تعیین مقدار بیت نهان‌نگاری شده با توجه به کلمه یا عبارت بکار گرفته‌شده در متن با استفاده از دیکشنری کلمات و عبارات مترادف و ارزش قراردادی که برای هر یک از دو کلمه یا عبارت مترادف در نظر گرفته‌شده است .

۲-۳-۲-۷ استفاده از برخی کاراکترهای خاص در متن (در ابتدای پاراگراف و ...)

در این روش، تعدادی کاراکتر خاص از کلمات مشخص به‌عنوان محلی برای مخفی کردن اطلاعات انتخاب می‌شود . در ساده‌ترین فرم اگر اولین کاراکتر از اولین کلمه هر پاراگراف انتخاب شود در یک حالت به‌وسیله قرار دادن اولین کاراکترهای این کلمات در کنار هم اطلاعات مخفی‌شده استخراج می‌شود[۲۱]. به‌عنوان‌مثال با استفاده از اشعار قدیمی ایران می‌توان این کار را انجام داد . این روش احتیاج به قدرت ذهنی قوی دارد و زمان زیادی هم می‌برد . نیاز به متن خاص داشته و در همه متن‌ها نمی‌توان از این روش استفاده کرد.

۲-۳-۲-۸ املاء متفاوت (در دو زبان متفاوت)

در این روش از کلماتی که در زبان‌های مختلف دارای دو املاء متفاوت هستند جهت نهان‌نگاری استفاده می‌کنیم [۲۲]. برای نهان‌نگاری بیت ۱ از نگارش کلمه در زبان اول و برای نهان‌نگاری بیت صفر از نگارش کلمه در زبان دوم استفاده می‌کنیم .گیرنده پیام با در اختیار داشتن بانک اطلاعاتی لغت‌هایی که دارای دو املاء صحیح ولی متفاوت هستند ، می‌تواند عمل نهان کاوی را انجام دهد .

۲-۳-۲-۹ استفاده از کاراکتر مخفی در متن (کاراکتر zwnj[۳۰])

در استاندارد یونیکد ، کاراکترهایی وجود دارد که بدون طول هستند و از آن‌ها به‌عنوان ابزاری جهت نهان‌نگاری استفاده می‌شود . کاراکترهای zwnj و zwc[۳۱] نمونه‌هایی هستند که با استفاده از درج آن‌ها نسبت به موقعیت کاراکترهای خاص نگارشی (مانند کاما ، نقطه ، ویرگول ، فاصله و ...) در متن اصلی می‌توان با تغییر نامحسوس و مقاومت بالا متن پیامی را نهان‌نگاری کرد[۲۳] .

به‌عنوان‌مثال اگر بعد از کلمه و قبل از کاراکتر فاصله از کاراکتر مخفی zwnj استفاده شود بیت ۱ ذخیره می‌شود و اگر بعد از کلمه از این کاراکتر استفاده نشود بیت صفر ذخیره می‌شود .

۲-۳-۲-۱۰ کاراکتر فاصله

در تمامی زبان‌های دنیا برای جداسازی کلمات از کاراکتر فاصله استفاده می‌شود درنتیجه با استفاده صحیح از کاراکتر فاصله می‌توان حجم زیادی از اطلاعات را نهان‌نگاری کرد .با توجه به اینکه از کاراکتر فاصله تنها به‌عنوآن‌یک فضای خالی بین کلمات استفاده می‌شود ، بسیاری از تغییرات که ازلحاظ نگارشی بر روی این کاراکتر اعمال می‌شود نشان داده نشده و بنابراین این تغییرات از دید بیننده مخفی باقی می‌ماند .

30 Zero-width non-joiner
31 zero-width-character

فصل سوم : روش پیشنهادی

مقدمه

برقراری ارتباط به معنای ارسال و دریافت داده و اطلاعات مشخص (پیام) از طریق رسانه‌های مختلف هست که این اطلاعات می‌توانند به انواع مختلف صوت، تصویر و متون نوشتاری باشند. هدف از برقراری ارتباط درواقع ارسال اطلاعات خاصی (پیام) از طریق فرستنده به گیرنده می‌باشد. در این میان گاها نیاز است که این پیام فقط به دست گیرنده برسد و اصطلاحاً پیام فاش نشود اما با توجه به اینکه رسانه‌های مورداستفاده برای ارسال اطلاعات از قبیل اینترنت و تلفن به‌صورت عمومی می‌باشند لذا امکان سرقت رفتن و لو رفتن اطلاعات وجود دارد. بنابراین از همان ابتدا برای ارسال پیام‌های سری از روش‌های مختلفی برای ارسال مخفی استفاده‌شده است. مثلاً در ابتدا،پیام به‌صورت رمزنگاری‌شده و با استفاده از کدهای مشخص، ارسال می‌شده است. عیب این روش در این است که از ظاهر اطلاعات مشخص است که این اطلاعات رمزنگاری‌شده می‌باشد. بعداً برای رفع این نقص، اقدام به نهان کردن پیام در انواع مختلف اطلاعات از قبیل صوت، تصویر و متون نوشتاری شده است که در این حالت با توجه به اینکه وجود پیام مخفی در شکل ظاهری اطلاعات قابل‌تشخیص نیست لذا احتمال لو رفتن پیام بسیار کمتر می‌باشد. این عمل مخفی کردن یک پیام در اطلاعات دیگر به‌عنوان نهان‌نگاری شناخته‌شده است. سه فاکتور امنیت، ظرفیت و مقاومت در مقابل تغییرات جزئی برای بررسی کار آیی روش‌های مختلف نهان‌نگاری تعریف‌شده‌اند. در نهان‌نگاری تصویر، مخفی کردن اطلاعات به معنای تغییر در تعدادی از پیکسل‌های تصویر می‌باشد. ایجاد تغییرات کم در تصویر قابل‌تشخیص نمی‌باشد اما ایجاد تغییرات بیشتر به‌راحتی قابل‌تشخیص خواهد بود لذا نهان‌نگاری تصویر از ظرفیت کمی برخوردار است. در ارتباط با نهان‌نگاری صوت نیز همین مشکل وجود دارد. نهان‌نگاری متن به معنای ایجاد تغییرات در دو شکل ظاهری و معنایی متن می‌باشد. تغییرات ظاهری شامل تغییر ویژگی‌های مختلف کاراکترها، موقعیت کاراکترها، کلمات و خطوط و غیره می‌باشد. تغییرات معنایی نیز به معنی استفاده از قابلیت‌های زبانی برای جایگزینی تعدادی از کلمات با کلمات دیگر مانند مترادف‌ها می‌باشد. با توجه به تنوع زیاد کاراکترها و کلمات موجود در متن، لذا امکان

ذخیره‌سازی اطلاعات بیشتری در این روش نسبت به دو روش دیگر نهان‌نگاری وجود دارد و این یعنی ظرفیت نهان‌نگاری متن بالا می‌باشد. در این فصل نهان‌نگاری متن به روش جدید ارائه خواهد شد.

۳ نهان‌نگاری متون فارسی

نهان‌نگاری متن به معنای ایجاد تغییرات در دو شکل ظاهری و معنایی متن می‌باشد. تغییرات ظاهری شامل تغییر ویژگی‌های مختلف کاراکترها، موقعیت کاراکترها، کلمات و خطوط و غیره می‌باشد. تغییرات معنایی نیز به معنی استفاده از قابلیت‌های زبانی برای جایگزینی تعدادی از کلمات با کلمات دیگر مانند مترادف‌ها می‌باشد. با توجه به تنوع زیاد کاراکترها و کلمات موجود در متن، لذا امکان ذخیره‌سازی اطلاعات زیادی در این روش نهان‌نگاری وجود دارد و این یعنی ظرفیت نهان‌نگاری متن بالا می‌باشد.

برای نهان‌نگاری یک پیام در یک متن، ابتدا باید پیام به‌صورت کدهای باینری تبدیل شود و سپس بیت‌های صفر و یک به‌صورت تغییر مشخصه‌های ظاهری در متن ذخیره شوند. عمل تبدیل پیام به کدهای باینری به‌عنوان رمزنگاری شناخته‌شده است. که این عمل رمزنگاری بر اساس کدهای شناخته‌شده یونیکد یا اسکی و یا کدهای تعریف‌شده مشخصی انجام می‌شود. در مرحله آخر پیام نهان شده در متن باید استخراج شود که این مرحله به‌عنوان کشف پیام شناخته می‌شود. در مرحله کشف پیام ابتدا بیت‌های باینری استخراج و بر اساس الگوریتم رمزنگاری به پیام اصلی تبدیل می‌شوند. در این پایان‌نامه هم بر روی الگوریتم جدید برای رمزنگاری و هم الگوریتم جدید نهان‌نگاری متون فارسی کارشده است. در این فصل ابتدا رمزنگاری پیام تشریح می‌شود و سپس به نهان‌نگاری پیام در متن پرداخته می‌شود.

۳-۱ رمزنگاری

قبل از نهان‌نگاری، پیام موردنظر ابتدا به کدهای تعریف‌شده تبدیل می‌شود که درواقع این کدها برای کاراکترهای مختلف پیام، کد باینری مشخص تعریف می‌کند. اولین کد استاندارد برای تخصیص کاراکترها، کد اسکی (ASCII) بوده است که شامل ۱۲۸ کاراکتر به زبان انگلیسی می‌باشد و به هر کاراکتر یک عدد از 0 تا 127 اختصاص می‌دهد که درواقع هر کاراکتر در یک‌رشته بیت ۷ تایی می‌گیرد. با توجه به تعداد زبان‌ها و نیاز به یک کد استاندارد برای همه کاراکترها، کد استاندارد یونیک تعریف‌شده است که بیش از 65535 کاراکتر در آن تعریف‌شده است و هر کاراکتر یک‌رشته بیت ۱۶ تایی می‌گیرد. لذا برای نهان‌نگاری یک پیام، ابتدا باید بر اساس کد استاندارد مثلاً یونیک، کاراکترهای پیام به رشته بیت‌های مشخص تبدیل شوند.

البته حضور رشته بیت‌های ۱۶ تایی در استاندارد یونیک برای هر کاراکتر باعث افزایش طول پیام می‌شود و در آن صورت برای نهان کردن یک پیام کوتاه نیاز به متون بلند دارد. در اینجا با توجه به اینکه قرار است یک پیام به زبان فارسی به در متن فارسی نهان شود لذا تنها از ۳۲ حرف فارسی به‌علاوه اعداد 9-0 و کاراکترهای انفصال موردنیاز (ویرگول، دونقطه، نقطه سرخط، نقطه‌ویرگول، علامت سؤال، خط تیره، علامت تعجب و علامت درصد) استفاده می‌شود که مجموعاً ۵۰ کاراکتر می‌باشد لذا می‌توان با رشته بیت‌های ۶ تایی این کاراکترها را رمزنگاری نمود.

در یک حالت خاص ممکن است متن پیام فقط شامل حروف و کاراکترهای انفصال باشد (اعداد نداشته باشد)، در این صورت تعداد کاراکترها ۴۰ خواهد شد که بازهم به ۶ بیت برای هر کاراکتر نیاز است. البته این حالت خاص در مرحله فشرده‌سازی به کمک مدل‌های زبانی می‌تواند به کاهش بیشتر تعداد بیت‌ها کمک کند. این مدل‌ها در ادامه این بخش بررسی خواهند شد.

در جدول (۳-۱) تفاوت طول رشته بیت برای کدگذاری با استاندارد یونیکد و روش پیشنهادی برای یک پیام خاص آورده شده است.

همان‌طور که در جدول آورده شده است در این روش بیش از ۶۰ درصد کاهش در طول رشته بیت‌ها خواهیم داشت.

علاوه براین برای کاهش بیشتر طول پیام از تکنیک‌های تحت عنوان فشرده‌سازی استفاده می‌شود که این تکنیک‌ها از مدل‌های زبانی استفاده می‌کند بدین‌صورت که در هر زبانی، فرکانس تکرار بعضی از حروف بیشتر است و یا احتمال وقوع تعدادی از حروف به دنبال حروف مشخصی بیشتر از سایر حروف است. این مدل‌های زبانی در ادامه بررسی خواهند شد.

جدول ۳. ۶ مقایسه طول رشته بیت در کد یونیکد و کدهای انتخابی

پیام	تعداد کاراکتر	طول رشته بیت با استاندارد یونیکد	طول رشته بیت با استفاده از روش پیشنهادی	تفاوت طول رشته بیت‌ها	درصد کاهش طول پیام
هوا سرد است	۱۴	۲۲۴	۸۴	۱۴۰	۶۲.۵
جلسه رأس ساعت ۹ تشکیل می‌شود	۳۲	۵۱۲	۱۹۲	۳۲۰	۶۲.۵
تا فردا می‌آیم	۱۷	۲۷۲	۱۰۲	۱۷۰	۶۲.۵

مدل یونیگرام[۳۲]

این مدل درواقع فرکانس تکرار حروف مختلف را در یک‌زبان مشخص می‌کند این‌که کدام حرف‌ها بیشتر تکرار خواهند شد. در هر زبانی با توجه به ویژگی‌های تکلم و ساختار نگارشی آن ، تعدادی از حروف بیشتر از سایرین کاربرد دارند مثلاً در زبان فارسی حروف "الف" یا "ر" حروف پرکاربردی هستند که در هم در اول، وسط و یا آخر کلمات می‌توانند حضورداشته باشند درحالی‌که حروفی مانند "پ"، "ژ"، "چ" و غیره به دو دلیل کمتر در زبان فارسی استفاده می‌شوند دلیل اول آن به‌سختی تکلم این حروف در کنار حروف دیگر می‌باشد بدین‌صورت که تنها با تعدادی محدود حروف دیگر هم‌نشین می‌شوند و دلیل دوم اینکه با توجه به اینکه در زبان فارسی کلمات عاریتی از زبان عربی بسیار زیاد می‌باشد و این دسته حروف نیز در زبان عربی حضور ندارند لذا استفاده آن در زبان فارسی محدود می‌باشد. لذا برای استخراج حروف با فرکانس تکرار بالا، متون متنوعی از سبک‌های مختلف شامل علمی، ادبی، اقتصادی، سیاسی بررسی‌شده‌اند و تعداد تکرار حروف مختلف در این متون بررسی‌شده و سپس نتایج آن‌ها در کنار هم قرار داده‌شده‌اند و نهایتاً با توجه به نتایج به‌دست‌آمده مشخص شد که تعدادی مشخص از حروف، بسیار بیشتر از حروف دیگر، در انواع مختلف متون، مورداستفاده قرار می‌گیرند. این حروف عبارت‌اند از " ا ی ر د ن و ه ت ب م". علاوه بر این حروف کاراکتر فاصله که برای جداسازی کلمات استفاده می‌شود دارای بیشترین تکرار می‌باشد. ازآنجایی‌که هدف فشرده‌سازی رمزنگاری برای کاهش تعداد بیت‌های لازم می‌باشد لذا در اینجا ۸ کاراکتر که دارای بیشترین تکرار می‌باشند، انتخاب‌شده‌اند و در یک گروه جدا قرار داده می‌شوند. این ۸ کاراکتر شامل "فاصله ا ی ر د ن و ه" می‌باشند. در جدول (۳-۲) فرکانس تکرار این حروف برای متون

[32] Unigram

مختلف و مقایسه آن‌ها با تعداد کل کاراکترها آورده شده است. همان‌طور که در جدول زیر دیده می‌شود برای انواع مختلف متن‌ها، این گروه ۸ تایی در حدود ۶۰ درصد کل کاراکترهای متن را تشکیل می‌دهند.

جدول ۳. ۷مقایسه فرکانس تکرار کاراکترهای خاص در مقابل کل کاراکترهای متون مختلف

متن	تعداد کل کاراکترها	تعداد کاراکترهای خاص								مجموع کاراکترهای خاص	درصد کاراکترهای خاص به کاراکترها
		فاصله	ا	ی	ر	د	ن	و	ه		
ورزشی	۸۹۲۹	۱۸۵۶	۹۲۱	۸۲۴	۴۶۰	۵۵۵	۴۵۰	۴۱۸	۳۶۵	۵۸۴۹	۶۵
اجتماعی	۸۰۵۰	۱۶۳۵	۹۴۱	۶۴۷	۴۸۴	۴۵۴	۴۳۲	۳۳۱	۴۴۵	۵۳۶۹	۶۶
سیاسی	۳۲۰۷	۶۶۳	۴۰۲	۲۱۶	۱۸۲	۱۷۰	۱۸۱	۱۶۴	۱۸۰	۲۱۵۸	۶۷
روانشناسی	۵۵۷۸	۱۱۴۵	۶۵۲	۴۵۲	۳۷۱	۳۳۳	۲۹۷	۲۱۰	۲۷۵	۳۷۳۵	۶۶
خانوادگی	۷۷۸۱	۱۴۶۸	۹۶۲	۶۱۱	۴۵۶	۴۹۸	۴۸۴	۲۶۴	۳۹۵	۵۱۳۸	۶۶
تاریخی	۳۲۵۴	۷۰۸	۳۴۹	۲۴۶	۲۰۱	۲۱۰	۱۷۱	۱۴۳	۱۴۰	۲۱۶۸	۶۶
اجتماعی	۱۸۰۴	۳۴۹	۱۶۷	۱۳۴	۱۰۸	۹۶	۹۶	۸۳	۷۷	۱۱۱۰	۶۱
اقتصادی	۴۴۷۶	۹۲۴	۵۳۶	۳۱۹	۲۶۹	۲۴۴	۲۳۰	۲۰۹	۱۷۶	۲۹۰۷	۶۴
پزشکی	۱۱۰۴	۲۱۸	۱۰۰	۹۲	۶۱	۶۴	۵۰	۴۴	۴۷	۶۷۶	۶۱
فناوری	۷۱۶۶	۱۴۱۰	۷۵۷	۵۰۷	۴۴۷	۴۴۵	۳۸۱	۳۱۳	۳۲۸	۴۵۸۸	۶۴

به‌منظور کاهش تعداد بیت‌های پیام در مرحله رمزنگاری، می‌توان برای این گروه خاص از سه بیت استفاده نمود و برای بقیه کاراکترها از همان ۶ بیت که قبلاً توضیح داده شد، استفاده نمود. البته

به‌منظور تفکیک کاراکترها در مرحله بازیابی اطلاعات در سمت گیرنده، به یک بیت کنترلی نیاز می‌باشد تا در شروع هر کاراکتر تعیین کند که آیا این کاراکتر از دسته کاراکترهای خاص یا از دسته دیگر کاراکترها می‌باشد. مثلاً می‌توان فرض نمود که اگر کاراکتر از گروه خاص باشد آنگاه در مرحله رمزنگاری بیت اول ۱ اختصاص داده شود و در غیر این صورت بیت اول صفر باشد. بنابراین با این کار، برای هر کاراکتر خاص نیاز به ۴ بیت و هر کاراکتر غیر خاص نیاز به ۷ کاراکتر می‌باشد. البته درصورتی‌که متن پیام فقط شامل حروف فارسی و کاراکترهای ربطی باشد در آن صورت بعد از جداسازی گروه خاص از مجموعه کاراکترها، تعداد کاراکترهای غیر خاص برابر ۳۲ خواهد شد لذا برای تبدیل باینری آن‌ها تنها به ۵ بیت نیاز می‌باشد که به‌اضافه یک بیت کنترلی، هر کاراکتر غیر خاص به ۶ بیت نیاز دارد.

در اینجا می‌توان کاهش تعداد بیت‌های موردنیاز را با این روش فشرده‌سازی برای یکی از متون جدول بالا (مثلاً متن اول) محاسبه نمود.

حالت اول- اگر متن پیام فقط شامل حروف و کاراکترهای انفصال باشد

متن اول: تعداد کل کاراکترها = ۸۹۲۹، تعداد کاراکترهای خاص= ۵۸۴۹، تعداد کاراکترهای غیر خاص= ۳۰۸۰

تعداد بیت‌های موردنیاز بدون فشرده‌سازی یونیگرام= ۸۹۲۹*۶= ۵۳۵۷۴

تعداد بیت‌های موردنیاز با فشرده‌سازی یونیگرام= ۵۸۴۹*۴+۳۰۸۰*۶= ۴۱۸۷۶

درصد کاهش بیت موردنیاز= ۲۱ درصد

حالت دوم- اگر متن پیام شامل حروف، کاراکترهای انفصال و اعداد باشد

متن اول: تعداد کل کاراکترها = ۸۹۲۹، تعداد کاراکترهای خاص= ۵۸۴۹، تعداد کاراکترهای غیر خاص= ۳۰۸۰

تعداد بیت‌های موردنیاز بدون فشرده‌سازی یونیگرام= ۸۹۲۹*۶= ۵۳۵۷۴

تعداد بیت‌های موردنیاز با فشرده‌سازی یونیگرام= ۵۸۴۹*۴+۳۰۸۰*۷= ۴۴۹۵۶

درصد کاهش بیت موردنیاز= ۱۶ درصد

این عملیات برای متون مختلف در جدول (۳-۳) نشان داده‌شده است. همان‌طور که در این جدول مشاهده می‌گردد با فشرده‌سازی به روش یونیگرام تقریباً ۲۰ درصد کاهش در تعداد بیت‌های موردنیاز خواهیم داشت.

جدول ۳. ۸ میزان کاهش ظرفیت رمزنگاری به کمک مدل زبانی یونیگرام

متن	تعداد کل بیت‌ها بدون فشرده‌سازی	تعداد کل بیت‌ها با فشرده‌سازی		درصد کاهش بیت موردنیاز
		پیام شامل حروف و کاراکترهای انفصال	پیام شامل حروف و اعداد و کاراکترهای انفصال	
ورزشی	۵۳۵۷۴	۴۱۸۷۶	۴۴۹۵۶	۲۱
اجتماعی	۴۸۳۰۰	۴۰۲۴۳	۳۷۵۶۲	۲۲
سیاسی	۱۹۲۴۲	۱۵۹۷۵	۱۴۹۲۶	۲۲
روانشناسی	۳۳۴۶۸	۲۷۸۴۱	۲۵۹۹۸	۲۲
خانوادگی	۴۶۶۸۶	۳۹۰۵۳	۳۶۴۱۰	۲۲
تاریخی	۱۹۵۲۰	۱۶۲۷۴	۱۵۱۸۸	۲۲
اجتماعی	۱۰۸۱۶	۹۲۹۸	۸۶۰۴	۲۰
اقتصادی	۲۶۸۵۶	۲۲۶۱۱	۲۱۰۴۲	۲۱
پزشکی	۶۶۲۴	۵۷۰۰	۵۲۷۲	۲۰
فناوری	۴۲۹۹۶	۳۶۳۹۸	۳۳۸۲۰	۲۱

مدل زبانی بایگرام[۳۳]

این مدل به امکان حضور و یا هم‌نشینی حروف در کنار هم دیگر در یک کلمه اشاره می‌کند. درواقع با توجه به ویژگی‌های زبانی در خواستگاه حروف، امکان تلفظ بعضی از حضور به دنبال هم وجود ندارد. مثلاً در زبان فارسی مجموعه حروف به دودسته صامت و مصوت تقسیم می‌شوند و هرکدام از این دسته‌ها خود به چندین دسته دیگر تقسیم می‌شوند این دسته‌بندی‌ها بر اساس مکانیسم تولید و اندام‌های شرکت‌کننده در تولید آن‌ها انجام می‌شود. بعضی از حروف دارای تلفظ روان و بعضی دیگر تلفظ آن‌ها سخت می‌باشد بنابراین با توجه به‌سختی تلفظ بعضی از حروف امکان هم‌نشینی را ندارند و یا با فرکانس تکرار کم حضور دارند. از طرفی دیگر حضور تعدادی از حروف در کنار هم با فرکانس تکرار بالا دیده می‌شود. لذا می‌توان از این ویژگی زبانی برای کاهش رمزنگاری استفاده نمود. در اینجا این مدل به دو متن مختلف اعمال‌شده است و نتیجه آن در جداول (۳-۴) و (۳-۵) آورده شده است.

جدول ۳. ۹ مدل زبانی بایگرام برای یک متن ورزشی نمونه (فرکانس تکرار حروف بعد از چند حرف نمونه "ا ب ت ی ر")

	ا	ب	پ	ت	ث	ج	چ	ح	خ	د	ذ	ر	ز	ژ	س	ش	ص	ض	ط	ظ	ع	غ	ف	ق	ک	گ	ل	م	ن	و	ه	ی	فاصله
ا	۰	۲۱	۰	۲۳	۰	۱	۰	۵	۱	۴۸	۱	۴۲	۶۷	۰	۲۵	۲۵	۰	۱	۲	۰	۱۴	۱	۲۳	۵	۵	۲	۴۵	۴۴	۱۱۲	۶	۲۰	۱۲۸	۹۳
ب	۴۳	۱	۱	۵	۰	۰	۰	۵	۲۱	۲۱	۱	۵۵	۷	۰	۱	۱	۰	۰	۲	۰	۱۷	۰	۰	۰	۱	۰	۵	۱	۵	۴	۷۲	۴۵	۲۱
ت	۵۵	۱	۰	۰	۱	۰	۲	۷	۰	۰	۱	۲۱	۰	۰	۲	۰	۰	۰	۰	۰	۲	۲	۱۰	۲	۰	۰	۲	۱۲	۱۴	۱۴	۱۲	۱۲	۸۵
ی	۱۲	۷	۱	۱۶	۰	۵	۰	۲	۱	۱۷۱	۱	۱۲	۱۲	۷	۲۰	۰	۰	۲	۰	۵	۰	۰	۰	۱۲	۱	۰	۱۲	۱۶	۵۸	۱	۴	۵۴	۲۵۲
ر	۷۲	۱۲	۱	۱۵	۰	۱	۲	۱	۲	۴۹	۰	۱۵	۱۵	۶	۲	۷	۲	۰	۰	۰	۰	۱	۱۵	۰	۷	۰	۲	۲	۲	۴۵	۲	۵۴	۱۲۵

جدول ۳. ۱۰ مدل زبانی بایگرام برای متن اجتماعی نمونه (تکرار حروف بعد از چند حرف نمونه "ا ب ت ی ر")

[33] Bigram

	ا	ب	پ	ت	ث	ج	چ	ح	خ	د	ذ	ر	ز	ژ	س	ش	ص	ض	ط	ظ	ع	غ	ف	ق	ک	گ	ل	م	ن	و	ه	ی	فاصله
ا	۰	۲۴	۰	۲۲	۰	۴	۰	۴	۶۶	۳۳	۱	۵۲	۶۴	۰	۵۶	۲۷	۹	۱	۳	۰	۲۰	۱	۱۸	۱۲	۶	۷	۳۶	۶۰	۱۵۴	۲۴	۴۰	۱۳۱	۱۳۶
ب	۶۳	۰	۱	۱۳	۰	۰	۰	۰	۰	۴	۰	۴۹	۰	۰	۱۳	۰	۰	۰	۴	۰	۱	۱	۱	۰	۱	۹	۶	۰	۵	۲۳	۵۸	۲۰	۲۰
ت	۲۵	۱۳	۰	۰	۱	۰	۰	۷	۵	۰	۰	۲۰	۰	۰	۱۱	۲	۲	۱	۰	۰	۷	۰	۶	۱	۷	۰	۲	۱۰	۱۴	۱۲	۲۸	۱۲	۷۸
ی	۵۴	۱	۰	۱۸	۰	۱۲	۵	۲	۱	۴۹	۰	۲۱	۱۲	۰	۱۳	۱۲	۰	۰	۱	۰	۵	۰	۴	۱	۲۰	۴	۱۰	۱۶	۵۸	۴	۶	۱۹	۲۹۲
ر	۹۸	۳	۰	۲۰	۰	۲	۱۱	۱	۴	۴۶	۰	۲	۱	۰	۲۹	۷	۱	۱	۰	۰	۵	۲	۷	۱۱	۴	۴	۰	۱	۱۰	۲۴	۴	۵۰	۱۵۴

همان‌طور که در جداول بالا دیده می‌شود تکرار بعضی از کاراکترها به دنبال حروف دیگر بیشتر از سایر حروف می‌باشد مثلاً در متن اول بعد از حرف الف کاراکترهای "د ر ز ل م ن ی فاصله س ش" بیشترین تکرار رادارند و در متن دوم بعد از حرف الف کاراکترهای "خ ر ز س م ن ی فاصله ل ه " دارای بیشترین تکرار هستند همان‌طور که دیده می‌شود کاراکترهای معینی بعد از الف بیشترین تکرار رادارند. بنابراین می‌توان کاراکترهای بعد از الف را به دودسته ۸ کاراکتر خاص و مابقی کاراکترهای غیر خاص تقسیم نمود و برای کاراکترهای خاص از ۴ بیت (یک بیت کنترلی و سه بیت شماره کاراکتر) و کاراکترهای غیر خاص از ۷ یا ۶ بیت (بسته به نوع پیام که آیا شامل اعداد می‌شود یا خیر) استفاده نمود.

این الگوریتم برای سایر حروف نیز قابل اجرا می‌باشد مثلاً در جداول بالا به‌صورت موردی مدل بایگرام برای حروف "ب ت ی ر" نیز نشان داده‌شده است همان‌طور که مشاهده می‌گردد برای این

حروف نیز دسته حروف خاص وجود دارد. بنابراین به کمک این الگوریتم می‌توان ظرفیت لازم رمزنگاری را کاهش داد.

از مقایسه دو مدل زبانی یونیگرام و بایگرام دیده می‌شود که الگوریتم رمزنگاری یونیگرام به‌مراتب ساده‌تر است و احتمال خطا کمتر می‌باشد. از طرف دیگر برای مقایسه میزان کاهش ظرفیت رمزنگاری، با توجه به اینکه الگوریتم بایگرام تمامی حروف را درگیر می‌کند لذا ممکن است در نگاه اول این‌طور به نظر برسد که میزان کاهش ظرفیت موردنیاز در این مدل بیشتر است اما باید توجه شود که فرکانس تکرار تعدادی از حروف (پ ث ج چ ح خ ذ ز ژ ص ض ط ظ ع غ ف ق ک گ)، تقریباً ۲۰حرف، بسیار کم می‌باشد. لذا مسئله تکرار حروف بعدازآن‌ها عملاً بی‌معنی می‌باشد بدین معنی که کمک زیادی به کاهش ظرفیت رمزنگاری نخواهند نمود. از طرف دیگر در این مدل، تعدادی از حالت‌های حروف پرتکرار نسبت به مدل یونیگرام کم می‌شوند. به‌عنوان‌مثال کاراکتر الف در دو مدل یونیگرام و بایگرام در اینجا مقایسه می‌شوند:

- در متن اول حرف الف به تعداد ۹۲۱ بار تکرار می‌شود لذا با مدل یونیگرام تعداد بیت‌های کاهش‌یافته به‌صورت زیر محاسبه می‌گردد:

  تعداد بیت‌های کاهش‌یافته = ۹۲۱*۲ = ۱۸۴۱ (هر کاراکتر الف به ۴ بیت نیاز دارد که در مقایسه با ۶ بیت دو بیت کاهش داریم).
- مدل بایگرام در متن اول: بعد از حرف الف، ۸ کاراکتر "د ر ز ل م ن ی فاصله" بیشترین تکرار رادارند که درمجموع ۶۸۴ بار این حروف تکرار می‌شوند. لذا با مدل یونیگرام تعداد بیت‌های کاهش‌یافته به‌صورت زیر محاسبه می‌گردد:

  تعداد بیت‌های کاهش‌یافته = ۶۸۴*۲ = ۱۳۶۸

بنابراین مشاهده می‌گردد که برای حروف پرتکرار مدل بایگرام نسبت به مدل یونیگرام کاهش ظرفیت کمتری دارد.

لذا در بایگرام تقریباً ۲۰ حرف کم تکرار (پ ث ج چ ح خ ذ ز ژ ص ض ط ظ ع غ ف ق ک گ) نقشی در کاهش ظرفیت ندارند و برای ۸ حرف پرتکرار (فاصله ا ی ر د ن و ه) نیز تأثیر کمتری نسبت به مدل یونیگرام دارد. در اینجا ۴ الی ۵ حرف(ب ت س ش ل) باقی می‌ماند که در مدل یونیگرام لحاظ نشده‌اند اما از فرکانس بالای متوسطی برخوردارند که این دسته حروف می‌توانند به بهبود وضعیت مدل بایگرام کمک بکنند. بدین معنی که این حروف می‌توانند خود منجر به کاهش ظرفیت رمزنگاری شوند.

درنهایت با در نظر گرفتن اثر حروف پرتکرار و حروف متوسط در دو مدل یونیگرام و بایگرام می‌توان نتیجه گرفت که تأثیر این دو مدل در کاهش ظرفیت موردنیاز رمزنگاری پیام تقریباً برابر می‌باشد. لذا با توجه سادگی الگوریتم یونیگرام از این مدل برای رمزنگاری در این پایان‌نامه استفاده می‌شود.

۳-۲ نهان‌نگاری و کشف پیام

نهان‌نگاری متن به معنای ایجاد تغییرات در دو شکل ظاهری و معنایی متن می‌باشد. تغییرات ظاهری شامل تغییر ویژگی‌های مختلف کاراکترها، موقعیت کاراکترها، کلمات و خطوط و غیره می-باشد. تغییرات معنایی نیز به معنی استفاده از قابلیت‌های زبانی برای جایگزینی تعدادی از کلمات با کلمات دیگر مانند مترادف‌ها می‌باشد.با توجه به محدودیت‌های جایگزینی کلمات درروش تغییرات معنایی،این روش ظرفیت نهان‌نگاری کمی دارد. لذا در این پایان‌نامه بر روی روش تغییرات ظاهری متن کارشده است.

کاراکترهای موجود در متن دارای ویژگی‌های مختلف نوشتاری از قبیل نوع قلم، سایز قلم، رنگ قلم، ضخامت قلم و ... می‌باشند. لذا می‌توان با تغییر در هرکدام از این ویژگی‌ها، اقدام به تغییر در متن نمود. مطالعاتی متنوعی بر روی تغییر ویژگی کاراکترها در زبان‌های مختلف صورت گرفته است. مثلاً در زبان فارسی با تغییر شکل تعدادی از کاراکترهای که شکل کشیده دارند یا تغییر شیب نوشتاری تعدادی از کاراکترها، نهان‌نگاری انجام‌شده است[۲۴] یا با استفاده از تغییر موقعیت مکانی خطوط و یا پاراگراف‌های متن اقدام به نهان‌نگاری شده است [۲۵] عیب این روش‌ها در این است که تغییرات از روی شکل ظاهری متن قابل‌تشخیص است. یکی از کاراکترهایی که تغییر در شکل ظاهری آن در متن قابل‌مشاهده نمی‌باشد کاراکتر فاصله می‌باشد لذا با تغییر ویژگی این کاراکتر پیام در متن ذخیره می‌شود. ازآنجایی‌که حضور این کاراکتر در متن زیاد می‌باشد لذا ظرفیت نهان‌نگاری این روش نسبتاً بالا می‌باشد[۲۶]. در این رساله بر روی کاراکتری دیگری که در زبان فارسی کاربرد دارد استفاده می‌شود این کاراکتر درواقع کاراکتر نیم‌فاصله می‌باشد. هرچند که این کاراکتر برای کلمات دوقسمتی (مانند می‌باشد)استفاده می‌شود اما در این رساله برای فاصله بین حروفی که به حرف قبلی یا بعدی نمی‌چسبند نیز از نیم‌فاصله استفاده می‌شود تا ظرفیت نهان‌نگاری بسیار بیشتر افزایش یابد. مثلاً در همان کلمه "می‌باشد" علاوه بر کاراکتر نیم‌فاصله موجود بین "می" و "باشد" می‌توان یک کاراکتر نیم‌فاصله نیز بین حروف "ا" و "ش" قرارداد. درنهایت برای افزایش ظرفیت نهان‌نگاری ترکیبی از دو کاراکتر نیم‌فاصله و کاراکتر فضای خالی استفاده می‌شود.

۳-۲-۱ نهان‌نگاری متن به کمک کاراکتر نیم‌فاصله

کاراکتر نیم‌فاصله (فاصله مجازی) در زبان فارسی برای جداسازی کلمات دوبخشی استفاده می‌شود این کاراکتر که در کد یونیکد با کد هگزا دسیمال 200c تعریف‌شده است تحت عنوان ZWNJ شناخته می‌شود .

هرچند که در زبان فارسی این کاراکتر برای کلمات دوقسمتی (مانند می‌باشد) استفاده می‌شود اما در این رساله برای فاصله بین حروفی که به حرف قبلی یا بعدی نمی‌چسبند نیز از نیم‌فاصله استفاده می‌شود تا ظرفیت نهان‌نگاری بسیار بیشتر افزایش یابد. مثلاً در همان کلمه "می‌باشد" علاوه بر کاراکتر نیم‌فاصله موجود بین "می" و "باشد" می‌توان یک کاراکتر نیم‌فاصله نیز بین حروف "ا" و "ش" قرارداد.. مثلاً در عبارت زیر می‌توان ۲۶ کاراکتر نیم‌فاصله قرارداد:

"نهان‌نگاری با استفاده از کاراکتر نیم‌فاصله در متون فارسی دارای ظرفیت بالایی می‌باشد."

همان‌طور که دیده می‌شود حضور کاراکتر نیم‌فاصله تغییری در شکل ظاهری متن ایجاد نمی‌کند.

اجرای نهان‌نگاری

پیام در مرحله رمزنگاری به رشته بیت‌های باینری صفر و یک تبدیل‌شده است. لذا این بیت‌ها باید در ویژگی کاراکترهای نیم‌فاصله ذخیره شود. ویژگی‌های کاراکتر شامل نوع قلم، سایز قلم، رنگ، ضخامت و خمیدگی قلم می‌باشد. هرکدام از این ویژگی‌ها به‌تنهایی و یا ترکیبی از آن‌ها می‌تواند با تغییر یا عدم‌تغییر نسبت به کاراکترهای مجاور، نشانگر حضور بیت یک یا بیت صفر باشند. در نهان‌نگاری علاوه بر خود پیام،در ابتدای متن طول پیام قرار داده می‌شود که این کار برای اطمینان از صحت پیام در مرحله کشف پیام انجام می‌شود. همچنین جهت جلوگیری از وقوع خطا، چندین بار عمل نهان‌نگاری در متن ایجاد می‌شود تا در مرحله کشف پیام با کنار هم گذاشتن پیام‌های کشف‌شده، خطاهای احتمالی حذف‌شده و امکان رسیدن به پیام صحیح بیشتر شود. باید دقت شود

که وقوع خطا در یک بیت می‌تواند منجر به از دست رفتن کل پیام شود. البته در ادامه این بخش برای جلوگیری از وقوع خطا علاوه بر تکرار پیاپی پیام در متن، از ترکیب ویژگی‌های مختلف استفاده خواهد شد. در ادامه ویژگی‌های مختلف کاراکتر نیم‌فاصله بررسی می‌شود و نحوه اجرای نهان‌نگاری و کشف پیام در مورد آن‌ها تشریح خواهد شد.

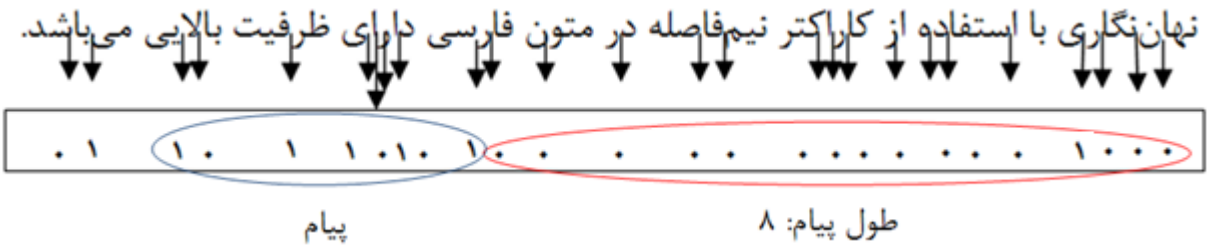
۳-۲-۱-۱ نوع قلم Font

برای نوشتن کاراکترها انواع مختلف شکل ظاهری وجود دارد که این اشکال به‌عنوان نوع قلم نوشتاری یا همان فونت شناخته‌شده هستند. در اینجا نهان‌نگاری و کشف پیام با استفاده از این ویژگی بررسی می‌شود.

نهان‌نگاری

در این حالت برای ذخیره بیت‌های صفر و یک در کاراکتر نیم‌فاصله نوع قلم کاراکتر نیم‌فاصله نسبت به کاراکتر قبلی را تغییر می‌دهیم. درصورتی‌که نوع قلم تغییر کرده باشد به مفهوم بیت یک و در صورت عدم‌تغییر به مفهوم بیت صفر می‌باشد. البته ۱۶ بیت ابتدایی برای ذخیره طول پیام استفاده می‌شود بدین‌صورت که طول پیام در ۱۶کاراکتر اول ذخیره می‌شود و سپس پیام شروع می‌شود. حضور طول پیام برای راستی آزمایی پیام در مرحله استخراج پیام از متن نهان‌نگاری شده، استفاده می‌شود. برای اطمینان از درستی پیام علاوه بر اینکه در ابتدای پیام طول پیام گذاشته می‌شود پیام چند بار در متن تکرار می‌شود. بعد از اتمام پیام دوباره نهان‌نگاری در قسمت بعدی متن شروع می‌شود. مثلاً در متن زیر پیام "10110101" نهان شده است. طول این پیام ۸ بیت می‌باشد لذا در ۱۶ بیت اول طول پیام ذخیره‌شده است بنابراین مشخصه ۸ کاراکتر بعدی برای نهان کردن پیام

استفاده می‌شود. اگر دقت شود در عبارت داده‌شده بعد از ۱۶ کاراکتر نیم‌فاصله ابتدایی، ۱۰ کاراکتر نیم‌فاصله دیگر وجود دارد اما با توجه به طول پیام که عدد ۸ می‌باشد لذا فقط ۸ کاراکتر بعدی برای نهان‌نگاری استفاده می‌شود و تغییر یا عدم‌تغییر دو کاراکتر بعدی تأثیری ندارد(شکل ۱.۳).



شکل(۳-۴) استفاده از کاراکتر نیم‌فاصله در نهان‌نگاری

در عبارت بالا فونت کلی به‌صورت "BNazanin" می‌باشد اما فونت تعدادی از کاراکترهای نیم‌فاصله به "Lotus" تغییر داده‌شده است. که این تغییر فونت در شکل ظاهری متن قابل‌تشخیص نمی‌باشد. شبه کد اجرای نهان‌نگاری با تغییر فونت در شکل (۳-۲) نشان داده‌شده است.

```
foreach(character of file)
{
        if(char==zwnj)
        {
                BitOfMessage=deleteFrontQueue();
                if(BitOfMessage==1)
                        ChangeFont(char);
                else
                        Don't ChangeFont(char);
        }
}
```

شکل (۳-۵) شبه کد اجرای نهان‌نگاری با تغییر فونت

استخراج پیام

برای استخراج پیام از متن نهان‌نگاری شده با شروع از ابتدای متن نوع قلم کاراکترهای نیم‌فاصله تعیین‌شده و تغییرات آن نسبت به کاراکتر قبلی آن بررسی می‌شود. وجود تغییر به معنی بیت یک و عدم‌تغییر به معنی بیت صفر می‌باشد. شبه کد کشف پیام از متن نهان‌نگاری شده با تغییر فونت در شکل (۳-۳) نشان داده‌شده است.

در مرحله بعد ۱۶ بیت اول برای طول پیام کنار گذاشته می‌شود و بقیه بیت‌ها برای کاراکترهای پیام استفاده می‌شود. درصورتی‌که مثلاً در مرحله رمزنگاری به هر کاراکتر پیام ۶ بیت نسبت داده‌شده باشد لذا در مرحله اکتشاف پیام، رشته بیت‌های استخراج‌شده به‌صورت ۶ تایی جداشده و بر اساس نحوه کدگذاری قراردادی اولیه، تبدیل به کاراکتر معادل می‌شوند و بدین‌صورت کل پیام استخراج می‌شود. و اگر رمزنگاری با استفاده از روش‌های فشرده‌سازی انجام‌شده باشد در آن صورت با توجه به بیت اول تعداد بیت‌ها دسته‌بندی می‌شوند مثلاً درروش مدل زبانی کاراکترهای پرتکرار ۴ بیت و مابقی کاراکترها ۷ بیت (پیام حاوی اعداد باشد)نیاز دارند. بنابراین با توجه به بیت اول، اگر مقدار آن‌یک باشد آنگاه از سه بیت بعد برای کاراکتر بعدی استفاده می‌شود و اگر صفر باشد آنگاه ۶ بیت بعدی به کاراکتر بعدی اختصاص داده می‌شود. مثلاً در شکل بالا با توجه به اینکه بعد از ۱۶ بیت اول، که به طول پیام اختصاص دارد، اولین بیت دارای مقدار یک می‌باشد پس باید سه بیت بعدازآن برای کاراکتر نوع پرتکرار استفاده شود و بیت بعدازاین ۳ بیت شروع کاراکتر بعدی خواهد بود. در پیام بالا دو کاراکتر از نوع کاراکترهای پرتکرار ذخیره‌شده‌اند. با فرض اینکه در مرحله رمزنگاری برای حروف پرتکرار ("فاصله ا ی ر د ن و ه")، به ترتیب اعداد ۰-۷ اختصاص داده‌شده باشد (یعنی به هرکدام سه بیت). آنگاه پیام بالا عبارت خواهد بود از: "اد".

```
foreach(character of file)
{
        if(char==zwnj)
        {
                if(Font is changed)

                        addFrontQueue(1);

                else
                        addFrontQueue(0);
        }
}
```

شکل (۳- ۶) شبه کد کشف پیام از متن نهان‌نگاری شده با تغییر فونت

۳-۲-۱-۲ سایز قلم

یکی دیگر از ویژگی‌های کاراکترها اندازه قلم نوشتاری می‌باشد که با توجه به اینکه خود کاراکتر نیم‌فاصله قابل‌رؤیت نیست لذا تغییر سایز آن نیز بر شکل متن تأثیری ندارد.

نهان‌نگاری

در این حالت برای ذخیره بیت‌های صفر و یک در کاراکتر نیم‌فاصله، سایز قلم کاراکتر نیم‌فاصله نسبت به کاراکتر قبلی را تغییر می‌دهیم. درصورتی‌که سایز قلم تغییر کرده باشد به مفهوم بیت یک و در صورت عدم‌تغییر به مفهوم بیت صفر می‌باشد البته ۱۶ بیت ابتدایی برای ذخیره طول پیام استفاده می‌شود بدین‌صورت که طول پیام در ۱۶کاراکتر اول ذخیره می‌شود و سپس پیام شروع می‌شود.

حضور طول پیام برای راستی آزمایی پیام در مرحله استخراج پیام از متن نهان‌نگاری شده استفاده می‌شود. برای اطمینان از درستی پیام علاوه بر اینکه در ابتدای پیام طول پیام گذاشته می‌شود پیام چند بار در متن تکرار می‌شود. بعد از اتمام پیام دوباره نهان‌نگاری در قسمت بعدی متن شروع می‌شود. شبه کد اجرای نهان‌نگاری با تغییر سایز قلم در شکل (۳-۴) نشان داده‌شده است.

```
foreach(character of file)
{
        if(char==zwnj)
        {
                BitOfMessage=deleteFrontQueue();
                if(BitOfMessage==1)
                        ChangeSize(char);
                else
                        Don't ChangeSize(char);
        }
}
```

شکل (۳-۴) شبه کد اجرای نهان‌نگاری با تغییر سایز قلم

استخراج پیام

برای استخراج پیام از متن نهان‌نگاری شده با شروع از ابتدای متن سایز قلم کاراکترهای نیم‌فاصله تعیین‌شده و تغییرات آن نسبت به کاراکتر قبلی آن بررسی می‌شود. وجود تغییر به معنی بیت یک و

عدم‌تغییر به معنی بیت صفر می‌باشد. شبه کد اجرای کشف پیام از متن نهان‌نگاری شده با تغییر سایز قلم در شکل (۳-۵) نشان داده‌شده است.

در مرحله ۱۶ بیت اول برای طول پیام کنار گذاشته می‌شود و بقیه بیت‌ها برای کاراکترهای پیام استفاده می‌شود. درصورتی‌که مثلاً در مرحله رمزنگاری به هر کاراکتر پیام ۶ بیت نسبت داده‌شده باشد لذا در مرحله اکتشاف پیام، رشته بیت‌های استخراج‌شده به‌صورت ۶ تایی جداشده و بر اساس نحوه کدگذاری قراردادی اولیه، تبدیل به کاراکتر معادل می‌شوند و بدین‌صورت کل پیام استخراج می‌شود. و اگر رمزنگاری با استفاده از روش‌های فشرده‌سازی انجام‌شده باشد در آن صورت با توجه به بیت اول تعداد بیت‌ها دسته‌بندی می‌شوند مثلاً درروش مدل زبانی کاراکترهای پرتکرار ۴ بیت و مابقی کاراکترها ۷ بیت (پیام حاوی اعداد باشد)نیاز دارند. بنابراین با توجه به بیت اول، اگر مقدار آن‌یک باشد آنگاه از سه بیت برای کاراکتر بعدی استفاده می‌شود و اگر صفر باشد آنگاه ۶ بیت بعدی به کاراکتر بعدی اختصاص داده می‌شود.

```
foreach(character of file)
{
        if(char==zwnj)
        {
                if(Size is changed)

                        addFrontQueue(1);

                else
                        addFrontQueue(0);
        }
}
```

شکل (۳-۵) شبه کد اجرای کشف پیام از متن نهان‌نگاری شده با تغییر سایز قلم

۳-۲-۱-۳ رنگ قلم

دیگر ویژگی کاراکترها رنگ قلم نوشتاری می‌باشد که با توجه به اینکه خود کاراکتر نیم‌فاصله قابل‌رؤیت نیست لذا تغییر سایز آن نیز بر شکل متن تأثیری ندارد.

نهان‌نگاری

در این حالت برای ذخیره بیت‌های صفر و یک در کاراکتر نیم‌فاصله رنگ قلم کاراکتر نیم‌فاصله نسبت به کاراکتر قبلی را تغییر می‌دهیم. درصورتی‌که رنگ قلم تغییر کرده باشد به مفهوم بیت یک و در صورت عدم‌تغییر به مفهوم بیت صفر می‌باشد. البته ۱۶ بیت ابتدایی برای ذخیره طول پیام استفاده می‌شود بدین‌صورت که طول پیام در ۱۶کاراکتر اول ذخیره می‌شود و سپس پیام شروع می-شود. حضور طول پیام برای راستی آزمایی پیام در مرحله استخراج پیام از متن نهان‌نگاری شده استفاده می‌شود. برای اطمینان از درستی پیام علاوه بر اینکه در ابتدای پیام طول پیام گذاشته می‌شود پیام چند بار در متن تکرار می‌شود. بعد از اتمام پیام دوباره نهان‌نگاری در قسمت بعدی متن شروع می‌شود. شبه کد اجرای نهان‌نگاری با تغییر رنگ قلم در شکل (۳-۶) نشان داده‌شده است.

```
foreach(character of file)
{
        if(char==zwnj)
        {
                BitOfMessage=deleteFrontQueue();
                if(BitOfMessage==1)
                        ChangeColor(char);
                else
                        Don't ChangeColor(char);
        }
}
```

شکل (۳-۶) شبه کد اجرای نهان‌نگاری با تغییر رنگ قلم

استخراج پیام

برای استخراج پیام از متن نهان‌نگاری شده با شروع از ابتدای متن رنگ قلم کاراکترهای نیم‌فاصله تعیین‌شده و تغییرات آن نسبت به کاراکتر قبلی آن بررسی می‌شود. وجود تغییر به معنی بیت یک و عدم‌تغییر به معنی بیت صفر می‌باشد. شبه کد اجرای کشف پیام از متن نهان‌نگاری شده با تغییر رنگ قلم در شکل (۳-۷) نشان داده‌شده است.

در مرحله بعد ۱۶ بیت اول برای طول پیام کنار گذاشته می‌شود و بقیه بیت‌ها برای کاراکترهای پیام استفاده می‌شود. درصورتی‌که مثلاً در مرحله رمزنگاری به هر کاراکتر پیام ۶ بیت نسبت داده‌شده باشد لذا در مرحله اکتشاف پیام، رشته بیت‌های استخراج‌شده به‌صورت ۶ تایی جداشده و بر اساس نحوه کدگذاری قراردادی اولیه، تبدیل به کاراکتر معادل می‌شوند و بدین‌صورت کل پیام استخراج می‌شود. و اگر رمزنگاری با استفاده از روش‌های فشرده‌سازی انجام‌شده باشد در آن صورت با توجه به

بیت اول تعداد بیت‌ها دسته‌بندی می‌شوند مثلاً درروش مدل زبانی کاراکترهای پرتکرار ۴ بیت و مابقی کاراکترها ۷ بیت (پیام حاوی اعداد باشد)نیاز دارند. بنابراین با توجه به بیت اول، اگر مقدار آن‌یک باشد آنگاه از سه بیت برای کاراکتر بعدی استفاده می‌شود و اگر صفر باشد آنگاه ۶ بیت بعدی به کاراکتر بعدی اختصاص داده می‌شود.

```
foreach(character of file)
{
        if(char==zwnj)
        {
                if(Color is changed)

                        addFrontQueue(1);

                else
                        addFrontQueue(0);
        }
}
```

شکل (۳-۷) شبه کد اجرای کشف پیام از متن نهان‌نگاری شده با تغییر رنگ قلم

۳-۲-۱-۴ ضخامت قلم (Bold)

ویژگی دیگر کاراکترها ضخامت قلم نوشتاری می‌باشد که با توجه به اینکه خود کاراکتر نیم‌فاصله قابل‌رؤیت نیست لذا تغییر سایز آن نیز بر شکل متن تأثیری ندارد.

نهان‌نگاری

در این حالت برای ذخیره بیت‌های صفر و یک در کاراکتر نیم‌فاصله ضخامت قلم کاراکتر نیم‌فاصله نسبت به کاراکتر قبلی را تغییر می‌دهیم. درصورتی‌که ضخامت قلم تغییر کرده باشد به مفهوم بیت یک و در صورت عدم‌تغییر به مفهوم بیت صفر می‌باشد. البته ۱۶ بیت ابتدایی برای ذخیره طول پیام استفاده می‌شود بدین‌صورت که طول پیام در ۱۶کاراکتر اول ذخیره می‌شود و سپس پیام شروع می-شود. حضور طول پیام برای راستی آزمایی پیام در مرحله استخراج پیام از متن نهان‌نگاری شده استفاده می‌شود. برای اطمینان از درستی پیام علاوه بر اینکه در ابتدای پیام طول پیام گذاشته می‌شود پیام چند بار در متن تکرار می‌شود. بعد از اتمام پیام دوباره نهان‌نگاری در قسمت بعدی متن شروع می‌شود. شبه کد اجرای نهان‌نگاری با تغییر ضخامت قلم در شکل (۳-۸) نشان داده‌شده است.

```
foreach(character of file)
{
        if(char==zwnj)
        {

                BitOfMessage=deleteFrontQueue();
                if(BitOfMessage==1)
                        ChangeBold(char);
                else
                        Don't ChangeBold(char);
        }
}
```

شکل (۳-۸) شبه کد اجرای نهان‌نگاری با تغییر ضخامت قلم

استخراج پیام

برای استخراج پیام از متن نهان‌نگاری شده با شروع از ابتدای متن ضخامت قلم کاراکترهای نیم‌فاصله تعیین‌شده و تغییرات آن نسبت به کاراکتر قبلی آن بررسی می‌شود. وجود تغییر به معنی بیت یک و عدم‌تغییر به معنی بیت صفر می‌باشد. شبه کد اجرای کشف پیام از متن نهان‌نگاری شده با تغییر ضخامت قلم در شکل (۳-۹) نشان داده‌شده است.

سپس ۱۶ بیت اول برای طول پیام کنار گذاشته می‌شود و بقیه بیت‌ها برای کاراکترهای پیام استفاده می‌شود. درصورتی‌که مثلاً در مرحله رمزنگاری به هر کاراکتر پیام ۶ بیت نسبت داده‌شده باشد لذا در مرحله اکتشاف پیام، رشته بیت‌های استخراج‌شده به‌صورت ۶ تایی جداشده و بر اساس نحوه کدگذاری قراردادی اولیه، تبدیل به کاراکتر معادل می‌شوند و بدین‌صورت کل پیام استخراج می‌شود. و اگر رمزنگاری با استفاده از روش‌های فشرده‌سازی انجام‌شده باشد در آن صورت با توجه به بیت اول تعداد بیت‌ها دسته‌بندی می‌شوند مثلاً درروش مدل زبانی کاراکترهای پرتکرار ۴ بیت و مابقی کاراکترها ۷ بیت (پیام حاوی اعداد باشد)نیاز دارند. بنابراین با توجه به بیت اول، اگر مقدار آن‌یک باشد آنگاه از سه بیت برای کاراکتر بعدی استفاده می‌شود و اگر صفر باشد آنگاه ۶ بیت بعدی به کاراکتر بعدی اختصاص داده می‌شود.

```
foreach(character of file)
{
        if(char==zwnj)
        {
                if(Bold is changed)

                        addFrontQueue(1);

                else
                        addFrontQueue(0);
        }
}
```

شکل (۳-۹) شبه کد اجرای کشف پیام از متن نهان‌نگاری شده با تغییر ضخامت قلم

۳-۲-۱-۵ خمیدگی قلم (Italic)

ویژگی دیگر کاراکترها خمیدگی قلم نوشتاری می‌باشد که با توجه به اینکه خود کاراکتر نیم‌فاصله قابل‌رؤیت نیست لذا تغییر سایز آن نیز بر شکل متن تأثیری ندارد.

نهان‌نگاری

در این حالت برای ذخیره بیت‌های صفر و یک در کاراکتر نیم‌فاصله خمیدگی قلم کاراکتر نیم‌فاصله نسبت به کاراکتر قبلی را تغییر می‌دهیم. درصورتی‌که خمیدگی قلم تغییر کرده باشد به مفهوم بیت یک و در صورت عدم‌تغییر به مفهوم بیت صفر می‌باشد. البته ۱۶ بیت ابتدایی برای ذخیره طول پیام استفاده می‌شود بدین‌صورت که طول پیام در ۱۶کاراکتر اول ذخیره می‌شود و سپس پیام شروع می‌شود. حضور طول پیام برای راستی آزمایی پیام در مرحله استخراج پیام از متن نهان‌نگاری

شده استفاده می‌شود. برای اطمینان از درستی پیام علاوه بر اینکه در ابتدای پیام طول پیام گذاشته می‌شود پیام چند بار در متن تکرار می‌شود. بعد از اتمام پیام دوباره نهان‌نگاری در قسمت بعدی متن شروع می‌شود. شبه کد اجرای نهان‌نگاری با تغییر خمیدگی قلم در شکل (۳-۱۰) نشان داده‌شده است.

```
foreach(character of file)
{
        if(char==zwnj)
        {

                BitOfMessage=deleteFrontQueue();
                if(BitOfMessage==1)
                        ChangeItalic(char);
                else
                        Don't ChangeItalic(char);
        }
}
```

شکل (۳-۱۰) شبه کد اجرای نهان‌نگاری با تغییر خمیدگی قلم

استخراج پیام

برای استخراج پیام از متن نهان‌نگاری شده با شروع از ابتدای متن خمیدگی قلم کاراکترهای نیم‌فاصله تعیین‌شده و تغییرات آن نسبت به کاراکتر قبلی آن بررسی می‌شود. وجود تغییر به معنی بیت یک و عدم‌تغییر به معنی بیت صفر می‌باشد. شبه کد اجرای کشف پیام از متن نهان‌نگاری شده با تغییر خمیدگی قلم در شکل (۳-۱۱) نشان داده‌شده است.

سپس ۱۶ بیت اول برای طول پیام کنار گذاشته می‌شود و بقیه بیت‌ها برای کاراکترهای پیام استفاده می‌شود. درصورتی‌که مثلاً در مرحله رمزنگاری به هر کاراکتر پیام ۶ بیت نسبت داده‌شده باشد لذا در مرحله اکتشاف پیام، رشته بیت‌های استخراج‌شده به‌صورت ۶ تایی جداشده و بر اساس نحوه کدگذاری قراردادی اولیه، تبدیل به کاراکتر معادل می‌شوند و بدین‌صورت کل پیام استخراج می‌شود. . و اگر رمزنگاری با استفاده از روش‌های فشرده‌سازی انجام‌شده باشد در آن صورت با توجه به بیت اول تعداد بیت‌ها دسته‌بندی می‌شوند مثلاً درروش مدل زبانی کاراکترهای پرتکرار ۴ بیت و مابقی کاراکترها ۷ بیت (پیام حاوی اعداد باشد)نیاز دارند. بنابراین با توجه به بیت اول، اگر مقدار آن‌یک باشد آنگاه از سه بیت برای کاراکتر بعدی استفاده می‌شود و اگر صفر باشد آنگاه ۶ بیت بعدی به کاراکتر بعدی اختصاص داده می‌شود.

```
foreach(character of file)
{
        if(char==zwnj)
        {
                if(Italic is changed)

                        addFrontQueue(1);

                else
                        addFrontQueue(0);
        }
}
```

شکل (۳-۱۱) شبه کد اجرای کشف پیام از متن نهان‌نگاری شده با تغییر خمیدگی قلم

۳-۲-۱-۶ ویژگی‌های ترکیبی

می‌توان از چندین ویژگی به‌صورت هم‌زمان استفاده نمود مثلاً دو یا چند ویژگی کاراکتر نیم‌فاصله را به‌صورت هم‌زمان تغییر دهیم. مزیت این روش به این است که می‌توان پیام‌های حاصل از هرکدام از ویژگی‌ها را استخراج نمود سپس با در کنار هم قرار دادن آن‌ها به پیام صحیح رسید. این روش وجود خطا در نهان‌نگاری را کاهش خواهد داد زیرا در صورت وجود خطا در یک ویژگی، مابقی ویژگی‌ها می‌توانند پیام درست را در خود جای دهند. لذا این روش باعث افزایش مقاومت نهان‌نگاری در مقابل تغییرات جزئی خواهد شد. شبه کد اجرای نهان‌نگاری با تغییر ویژگی‌های ترکیبی و کشف پیام از متن نهان‌نگاری شده با تغییر ویژگی‌های ترکیبی در شکل‌های (۳-۱۲) و (۳-۱۳)نشان داده‌شده است.

```
foreach(character of file)
{
        if(char==zwnj)
        {

        BitOfMessage=deleteFrontQueue();
                if(BitOfMessage==1)
                        ChangeFont&Size&Color&Bold&Italic(char);
                else
                        Don't ChangeFont&Size&Color&Bold&Italic(char);
        }
}
```

شکل (۳-۱۲) شبه کد اجرای نهان‌نگاری با تغییر ویژگی‌های ترکیبی

```
foreach(character of file)
{
        if(char==zwnj)
        {
                if(Font&Size&Color&Bold&Italic is changed)

                        addFrontQueue(1);

                else
                        addFrontQueue(0);
        }
}
```

شکل (۳-۱۳) شبه کد اجرای کشف پیام از متن نهان‌نگاری شده با تغییر ویژگی‌های ترکیبی

۳-۲-۲ نهان‌نگاری متن به کمک دو کاراکتر نیم‌فاصله و فاصله

برای افزایش بیشتر ظرفیت نهان‌نگاری متون فارسی می‌توان به‌صورت هم‌زمان از ویژگی‌های دو کاراکتر نیم‌فاصله و فاصله استفاده نمود. مثلاً در عبارت زیر علاوه بر ۲۶ کاراکتر نیم‌فاصله، ۱۲ کاراکتر فاصله نیز موجود می‌باشد. بنابراین درمجموع ۳۸ کاراکتر قابل‌استفاده برای نهان‌نگاری وجود دارد لذا ظرفیت این روش بیش از ۳ برابر ظرفیت استفاده تنها از کاراکتر فضای خالی می‌باشد.

"نهان‌نگاری با استفاده از کاراکتر نیم‌فاصله در متون فارسی دارای ظرفیت بالایی می‌باشد."

برای اجرای نهان‌نگاری مشابه حالت کاراکتر نیم‌فاصله، از همه ویژگی‌ها برای هر دو کاراکتر استفاده می‌شود. و برای کشف پیام نیز ویژگی‌های هر دو کاراکتر بررسی خواهد شد و مشابه بخش قبلی، ابتدا بیت‌ها استخراج و سپس کاراکترها کشف می‌شوند. شبه کد اجرای نهان‌نگاری با تغییر

ویژگی‌های ترکیبی برای کاراکترهای نیم‌فاصله و فاصله و کشف پیام از متن نهان‌نگاری شده با تغییر ویژگی‌های ترکیبی در شکل‌های (۳-۱۴) و (۳-۱۵)نشان داده‌شده است.

```
foreach(character of file)
{
        if(char==zwnj OR char=space)
        {

        BitOfMessage=deleteFrontQueue();
                if(BitOfMessage==1)
                        ChangeFont&Size&Color&Bold&Italic(char);
                else
                        Don't ChangeFont&Size&Color&Bold&Italic(char);
        }
}
```

شکل (۳-۱۴) شبه کد اجرای نهان‌نگاری با تغییر ویژگی‌های ترکیبی برای کاراکترهای نیم‌فاصله و فاصله

```
foreach(character of file)
{
        if(char==zwnj OR char=space)
        {
                if(Font&Size&Color&Bold&Italic is changed)

                        addFrontQueue(1);

                else
                        addFrontQueue(0);
        }
}
```

شکل (۳-۱۵) شبه کد اجرای کشف پیام از متن نهان‌نگاری شده با تغییر ویژگی‌های ترکیبی برای کاراکترهای نیم‌فاصله و فاصله

۳-۳ بررسی کار آیی روش نهان‌نگاری

کار آیی روش‌های نهان‌نگاری با استفاده از سه فاکتور ظرفیت، امنیت و مقاومت روش سنجیده می‌شود که در این بخش این سه فاکتور برای روش نهان‌نگاری پیشنهادی بررسی خواهند شد.

۳-۳-۱ ظرفیت نهان‌نگاری روش پیشنهادی

در این رساله برای فاصله بین حروفی که به حرف قبلی یا بعدی نمی‌چسبند نیز از نیم‌فاصله استفاده می‌شود تا ظرفیت نهان‌نگاری افزایش یابد. مثلاً در همان کلمه "می‌باشد" علاوه بر کاراکتر نیم‌فاصله موجود بین "می" و "باشد" می‌توان یک کاراکتر نیم‌فاصله نیز بین حروف "ا" و "ش" قرارداد. همچنین برای افزایش بیشتر ظرفیت نهان‌نگاری، علاوه بر کاراکتر نیم‌فاصله از کاراکتر فاصله (فضای خالی) نیز به‌صورت هم‌زمان استفاده می‌شود لذا ظرفیت نهان‌نگاری بسیار بیشتر افزایش

می‌یابد. مثلاً در عبارت زیر علاوه بر ۲۶ کاراکتر نیم‌فاصله، ۱۲ کاراکتر فاصله نیز وجود دارد که مجموعاً امکان ذخیره ۳۸ بیت را فراهم می‌کند. لذا در مقایسه باکارهای قبلی که تنها از کاراکتر فاصله استفاده‌شده است، ظرفیت نهان‌نگاری تقریباً سه برابر خواهد شد.

"نهان‌نگاری با استفاده از کاراکتر نیم‌فاصله در متون فارسی دارای ظرفیت بالایی می‌باشد."

۳-۳-۲ امنیت نهان‌نگاری روش پیشنهادی

از طرف دیگر، همان‌طور که در بخش‌های بالا ذکر شد تغییر هرکدام از ویژگی‌های کاراکتر نیم‌فاصله به معنای بیت یک و عدم‌تغییر آن به معنای بیت صفر لحاظ شده است. اما این نوع نهان‌نگاری ممکن است از امنیت بالایی برخوردار نباشد مثلاً یک شخص ثالث می‌تواند با دنبال کردن کاراکترهای متن متوجه تغییراتی در ویژگی بعضی از کاراکترها بشود بنابراین می‌تواند با دو بار وارسی متن به پیام نهان شده پی ببرد بدین‌صورت که یک‌بار تغییر ویژگی را به‌عنوان بیت یک و عدم‌تغییر را به‌عنوان بیت صفر و دفعه دوم بالعکس یعنی تغییر ویژگی را به‌عنوان بیت صفر و عدم‌تغییر را به‌عنوان بیت یک لحاظ کند.

برای رفع این مشکل می‌توان از الگوریتم مختلف استفاده کرد مثلاً قرارداد شود ۳۲ حرف زبان فارسی به دودسته اول (۱۶ حرف اول) و دوم (۱۶ حرف دوم) تقسیم شود سپس بر اساس کاراکتر قبلی کاراکتر نیم‌فاصله تعیین شود که آیا تغییر ویژگی کاراکتر به‌عنوان بیت یک ذخیره شود یا بیت صفر. مثلاً اگر کاراکتر قبلی از دسته اول حروف باشد آنگاه تغییر ویژگی کاراکتر نیم‌فاصله به‌عنوان بیت یک و عدم‌تغییر به‌عنوان بیت صفر تعیین شود و اگر کاراکتر قبلی از دسته دوم حروف باشد آنگاه تغییر ویژگی کاراکتر نیم‌فاصله به‌عنوان بیت صفر و عدم‌تغییر به‌عنوان بیت یک تعیین شود. البته می‌توان به‌جای دودسته به چند دسته تقسیم کرد و برای هر دسته یک حالت انتخاب شود.

بدین‌صورت در این رساله امنیت نهان‌نگاری خیلی بیشتر افزایش می‌یابد. شبه کدهای اجرای نهان‌نگاری و کشف پیام با تغییر الگوی نهان‌نگاری جهت افزایش امنیت نهان‌نگاری در شکل‌های (۳-۱۶) و (۳-۱۷) نشان داده‌شده است.

```
foreach(character of file)
{
        if(char==zwnj)
        {
                BitOfMessage=deleteFrontQueue();
                If(prior char is from set1)
                        if(BitOfMessage==1)
                                ChangeFont&Size&Color&Bold&Italic(char);
                        else
                                Don't ChangeFont&Size&Color&Bold&Italic(char);
                else
                        if(BitOfMessage==0)
                                ChangeFont&Size&Color&Bold&Italic(char);
                        else
                                Don't ChangeFont&Size&Color&Bold&Italic(char);
        }
}
```

شکل (۳-۱۶) شبه کد اجرای نهان‌نگاری با تغییر الگو جهت افزایش امنیت

```
foreach(character of file)
{
        if(char==zwnj)
        {
                If(prior char is from set1)
                        if(Font&Size&Color&Bold&Italic is changed)

                                addFrontQueue(1);
                        else
                                addFrontQueue(0);
                else
                        if(Font&Size&Color&Bold&Italic is changed)

                                addFrontQueue(0);
                        else
                                addFrontQueue(1);
        }
}
```

شکل (۳-۱۷) شبه کد اجرای کشف پیام با تغییر الگو جهت افزایش امنیت

۳-۳-۳ مقاومت نهان‌نگاری روش پیشنهادی

از طرف دیگر امکان وجود خطا در این شیوه نهان‌نگاری با تغییر ویژگی کاراکتر وجود دارد یعنی ممکن است سهواً یک بیت اشتباه پنهان شود مثلاً در جاییکه ویژگی کاراکتر باید تغییر داده شود این عمل انجام نشود. که این عمل باعث ایجاد خطا در زمان اکتشاف پیام از متن نهان‌نگاری شده می‌شود که در حالتی که برای رمزنگاری از روش‌های فشرده‌سازی استفاده‌شده باشد، که در آن طول هر کاراکتر دقیقاً مشخص نیست، منجر به از دست رفتن کامل پیام می‌شود. لذا باید نهان‌نگاری در مقابل تغییرات احتمالی کوچک پیام را همچنان حفظ بکند که به این ویژگی نهان‌نگاری، مقاومت روش

نهان‌نگاری گفته می‌شود. در این رساله برای افزایش مقاومت نهان‌نگاری ابتدا پیام را چند بار در متن تکرار می‌کنیم بدین‌صورت که درست بعد از اتمام پیام دوباره شروع به نهان نمودن پیام در ادامه متن برای چند بار می‌شود. تا در مرحله اکتشاف چندین بار پیام استخراج شود و نتایج باهم مقایسه شوند تا در صورت وجود خطا، به پیام درست دست‌یافته شود که این باعث افزایش مقاومت نهان‌نگاری می‌شود. علاوه بر این با ترکیب ویژگی‌های مختلف کاراکتر نیم‌فاصله (نوع، سایز، رنگ، ضخامت و خمیدگی قلم)، مقاومت نهان‌نگاری افزایش داده می‌شود در این حالت نتایج اکتشاف برای تمام ویژگی‌ها کنار هم گذاشته می‌شود تا به پیام درست دست‌یافت. لذا با این دو روش مقاومت نهان‌نگاری افزایش داده می‌شود.

با توجه به مطالب ذکرشده، نوآوری‌های این رساله افزایش ظرفیت و امنیت نهان‌نگاری متون فارسی می‌باشد. برای نهان‌نگاری از کاراکتر نیم‌فاصله به همراه کاراکتر فاصله استفاده‌شده است. همچنین با تغییر الگوی نهان‌نگاری اینکه تغییر یا عدم‌تغییر ویژگی به معنای صفر باشد یا یک، امنیت روش نهان‌نگاری افزایش‌یافته است. در فصل بعد با استفاده از روش پیشنهادی، عمل نهان‌نگاری برای متون مختلف انجام می‌شود و فاکتورهای نهان‌نگاری این روش با روش‌های قبلی مقایسه خواهد شد.

فصل چهارم : نتایج نهان‌نگاری

مقدمه

در این فصل چگونگی اجرای مراحل مختلف نهان‌نگاری به روش پیشنهادی برای چندین متن مختلف ارائه‌شده است. در ادامه فاکتورهای ارزیابی نهان‌نگاری از قبیل ظرفیت، امنیت و مقاومت برای روش پیشنهادی بررسی‌شده و سپس با سایر روش‌ها مقایسه می‌گردد. در ابتدا رمزنگاری پیام به روش پیشنهادی تشریح خواهد شد که این بخش با استفاده از روش‌های ذکرشده در فصل قبل اجرا خواهد شد. سپس در مرحله بعد نهان‌نگاری پیام در متن با روش پیشنهادی بررسی خواهد شد و بعدازآن به کشف پیام از متن نهان‌نگاری شده پرداخته خواهد شد. و در انتهای این فصل پارامترهای ارزیابی روش پیشنهادی بررسی‌شده و با سایر روش‌ها مقایسه می‌گردد.

۴-۱منابع مورداستفاده

در این پایان‌نامه روش نهان‌نگاری پیشنهادی بر روی چندین نوع مختلف متن از قبیل ادبی، علمی، ورزشی، روانشناسی، .. اجرا می‌شود. این متون از سایت‌های تحلیلی خبری عصر ایران و تابناک انتخاب‌شده‌اند. که لینک آن‌ها در جدول زیر آورده شده است.

جدول۴. ۱ لینک متون استفاده‌شده در پایان‌نامه

متن	لینک
ورزشی	http://www.asriran.com/fa/news/437536/%D8%B1%D8%A7%D8%B2-%D8%AA%D9%86%D8%A7%D8%B3%D8%A8-%D8%A7%D9%86%D8%AF%D8%A7%D9%85-%D9%BE%D8%B3-%D8%A7%D8%B2-40-%D8%B3%D8%A7%D9%84%DA%AF%DB%8C
اجتماعی	http://www.tabnak.ir/fa/news/547927/%D8%A8%D8%B1%DA%AF-%D8%A8%D8%B1%D9%86%D8%AF%D9%87-%D8%AF%D8%A7%D9%86%D8%B4-%D8%A2%D9%85%D9%88%D8%B2%D8%A7%D9%86-%D8%AF%D8%B1-%DA%A9%D9%86%DA%A9%D9%88%D8%B1-%DA%86%DB%8C%D8%B3%D8%AA
سیاسی	http://www.asriran.com/fa/news/433390/10-%D8%B1%D9%88%D8%B2-%D8%A8%D8%A7-%D8%AF%D8%A7%D8%B9%D8%B4-%D8%AF%D8%B1-%D8%AF%D9%84-%D8%AE%D9%84%D8%A7%D9%81%D8%AA-%D8%A7%D8%B3%D9%84%D8%A7%D9%85%D9%8A
روانشناسی	http://www.asriran.com/fa/news/432910/%D8%AF%D9%84%D8%A7%DB%8C%D9%84-%DA%AF%D8%B1%DB%8C%D9%87-%DA%A9%D8%B1%D8%AF%D9%86-%D8%A7%D9%81%D8%B1%D8%A7%D8%AF-%D8%AD%D8%B3%D8%A7%D8%B3
خانوادگی	http://www.asriran.com/fa/news/429025/%D8%A7%DB%8C%D9%86-10-%D8%B1%D8%A7%D8%A8%D8%B7%D9%87-%D8%B1%D8%A7-%D8%A8%D8%A7-%D8%B9%D8%B4%D9%82-%D8%A7%D8%B4%D8%AA%D8%A8%D8%A7%D9%87-%D9%86%DA%AF%DB%8C%D8%B1%DB%8C%D8%AF
تاریخی	http://www.asriran.com/fa/news/433455/%D8%AF%D9%88%D9%85-%D8%A2%D8%B0%D8%B1-%D8%A8%D8%A7-%DB%8C%D8%A7%D8%AF-%D8%B4%D8%B1%DB%8C%D8%B9%D8%AA%DB%8C-%D9%88-%D8%B3%D8%A7%D9%85%DB%8C-%D9%88-%D8%B3%D8%A7%D8%B9%D8%AF%DB%8C-%D9%88-%D8%B4%D8%A7%DB%8C%D8%AF-%D9%87%D9%85-%D8%AC%D9%84%D8%A7%D9%84
اجتماعی	http://www.asriran.com/fa/news/432885/%DA%86%D9%82%D8%AF%D8%B1-%D8%A8%D9%87-%D9%87%D9%85-

%D8%A8%D8%AF%D9%87%DA%A9%D8%A7%D8%B1%DB%8C%D9%85	
http://www.asriran.com/fa/news/433480/%D8%A7%D8%A8%D9%84%D8%A7%D8%BA-%D9%88%D8%A7%D9%85-60-%D9%85%DB%8C%D9%84%DB%8C%D9%88%D9%86%DB%8C-%D9%85%D8%B3%DA%A9%D9%86-%D8%B4%D8%B1%D8%A7%DB%8C%D8%B7-%D8%A7%D9%88%D8%B1%D8%A7%D9%82-%D9%88-%D8%A7%D9%82%D8%B3%D8%A7%D8%B7	اقتصادی
http://www.asriran.com/fa/news/433257/%DA%A9%D8%A7%D9%87%D8%B4-%DA%A9%D9%84%D8%B3%D8%AA%D8%B1%D9%88%D9%84-%D8%AE%D9%88%D9%86-%D8%A8%D8%A7-%DA%AF%D8%B1%D8%AF%D9%88	پزشکی
http://www.asriran.com/fa/news/422506/%DA%86%D8%A7%D9%84%D8%B4-%D8%A8%D8%B2%D8%B1%DA%AF-%D8%B3%DB%8C%D8%A7%D8%B3%D8%AA-%D8%AE%D8%A7%D8%B1%D8%AC%DB%8C-%D8%A7%DB%8C%D8%B1%D8%A7%D9%86	فناوری

۲-۴ رمزنگاری پیام

رمزنگاری پیام اولین مرحله در اجرای نهان‌نگاری پیام در متن می‌باشد در این مرحله پیام متنی به کدهای باینری تبدیل می‌شوند. اساس تبدیل کاراکترها به کدهای باینری استفاده از کدهای اسکی و یا یونیکد می‌باشد که کدهای استاندارد می‌باشند. کد اسکی که تنها برای حروف انگلیسی قابل‌استفاده می‌باشد اما کد یونیکد برای تمامی زبان‌های نوشتاری تعریف‌شده است با توجه به حضور تعداد بسیار زیاد کاراکترها در این کد، برای تبدیل کاراکتر به کد باینری به ۱۶ بیت برای هر کاراکتر نیاز می‌باشد. که این تعداد بیت موجب می‌شود که پیام طول رشته بیت بلندی داشته باشد و لذا برای نهان‌نگاری به متون بلندی نیاز دارند. اما با توجه به اینکه هدف این پروژه نهان نمودن پیام فارسی در متن فارسی می‌باشد لذا در یک پیام فقط کاراکترهای حروف الفبای فارسی، کاراکترهای انفصال و ارقام حضور دارند. با توجه به جدول (۴-۲) تعداد کل کاراکترها ۵۰ عدد می‌باشد بنابراین می‌توان با مرتب نمودن حروف الفبا و سایر کاراکترها، به هر کاراکتر یک عدد در فاصله صفرتا ۴۹ اختصاص داد و سپس کد باینری عدد مربوطه به‌عنوان کد کاراکتر تعیین شود.

جدول۴. ۲ کاراکترهای فارسی همراه با کد پیشنهادی آن‌ها (حروف الفبای فارسی، کاراکترهای انفصال و ارقام)

کاراکتر	ا	ب	پ	ت	ث	ج	چ	ح	خ	د	ذ	ر	ز	ژ	س	ش	ص	ض	ط	ظ	ع	غ	ف	ق	ک
شماره کاراکتر	۰	۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰	۱۱	۱۲	۱۳	۱۴	۱۵	۱۶	۱۷	۱۸	۱۹	۲۰	۲۱	۲۲	۲۳	۲۴
کد باینری	۰۰۰۰۰۰	۰۰۰۰۰۱	۰۰۰۰۱۰	۰۰۰۰۱۱	۰۰۰۱۰۰	۰۰۰۱۰۱	۰۰۰۱۱۰	۰۰۰۱۱۱	۰۰۱۰۰۰	۰۰۱۰۰۱	۰۰۱۰۱۰	۰۰۱۰۱۱	۰۰۱۱۰۰	۰۰۱۱۰۱	۰۰۱۱۱۰	۰۰۱۱۱۱	۰۱۰۰۰۰	۰۱۰۰۰۱	۰۱۰۰۱۰	۰۱۰۰۱۱	۰۱۰۱۰۰	۰۱۰۱۰۱	۰۱۰۱۱۰	۰۱۰۱۱۱	۰۱۱۰۰۰

کاراکتر	گ	ل	م	ن	و	ه	ی	فاصله	نیم‌فاصله	.	،	؛	:	؟	!	۰	۱	۲	۳	۴	۵	۶	۷	۸	۹
شماره کاراکتر	۲۵	۲۶	۲۷	۲۸	۲۹	۳۰	۳۱	۳۲	۳۳	۳۴	۳۵	۳۶	۳۷	۳۸	۳۹	۴۰	۴۱	۴۲	۴۳	۴۴	۴۵	۴۶	۴۷	۴۸	۴۹
کد باینری	۰۱۱۰۰۱	۰۱۱۰۱۰	۰۱۱۰۱۱	۰۱۱۱۰۰	۰۱۱۱۰۱	۰۱۱۱۱۰	۰۱۱۱۱۱	۱۰۰۰۰۰	۱۰۰۰۰۱	۱۰۰۰۱۰	۱۰۰۰۱۱	۱۰۰۱۰۰	۱۰۰۱۰۱	۱۰۰۱۱۰	۱۰۰۱۱۱	۱۰۱۰۰۰	۱۰۱۰۰۱	۱۰۱۰۱۰	۱۰۱۰۱۱	۱۰۱۱۰۰	۱۰۱۱۰۱	۱۰۱۱۱۰	۱۰۱۱۱۱	۱۱۰۰۰۰

بنابراین با این روش رمزنگاری به ازای هر کاراکتر به ۶ بیت نیاز می‌باشد. لذا تعداد بیت‌های موردنیاز برای رمز کردن پیام نسبت به استاندارد یونیکد خیلی کمتر است.

البته برای کاهش بیشتر طول رشته بیت‌های پیام رمز شده، می‌توان از مدل‌های زبانی استفاده نمود. همان‌طور که در فصل قبل ذکر شد یکی از مدل‌های زبانی که هم بیت‌های موردنیاز را کاهش می‌دهد و هم از پیچیدگی کمتری برخوردار است مدل یونیگرام می‌باشد. رمزنگاری کاراکترها با استفاده از مدل زبانی یونیگرام در جدول (۴-۳) نشان داده‌شده است.

جدول۴. ۳ کاراکترهای فارسی همراه با کد پیشنهادی آن‌ها با روش مدل زبانی یونیگرام (حروف الفبای فارسی، کاراکترهای انفصال و ارقام)

کاراکتر	فاصله	ا	د	ر	ن	و	ه	ی
شماره کاراکتر	۰	۱	۲	۳	۴	۵	۶	۷
کد باینری	۰۰۰	۰۰۱	۰۱۰	۰۱۱	۱۰۰	۱۰۱	۱۱۰	۱۱۱

کاراکتر	ب	پ	ت	ث	ج	چ	ح	خ	ذ	ز	ژ	س	ش	ص	ض	ط	ظ	ع	غ	ف	ق
شماره کاراکتر	۰	۸	۹	۱۰	۱۱	۱۲	۱۳	۱۴	۱۵	۱۶	۱۷	۱۸	۱۹	۲۰	۲۱	۲۲	۲۳	۲۴	۲۵	۲۶	۲۷
کد باینری	۰۰۰۰۰	۰۰۰۰۱	۰۰۰۱۰	۰۰۰۱۱	۰۰۱۰۰	۰۰۱۰۱	۰۰۱۱۰	۰۰۱۱۱	۰۱۰۰۰	۰۱۰۰۱	۰۱۰۱۰	۰۱۰۱۱	۰۱۱۰۰	۰۱۱۰۱	۰۱۱۱۰	۰۱۱۱۱	۱۰۰۰۰	۱۰۰۰۱	۱۰۰۱۰	۱۰۰۱۱	۱۰۱۰۰

کاراکتر	ک	گ	ل	م	نیم‌فاصله	.	،	؛	:	؟	!	۰	۱	۲	۳	۴	۵	۶	۷	۸	۹
شماره کاراکتر	۲۸	۲۹	۳۰	۳۱	۳۲	۳۳	۳۴	۳۵	۳۶	۳۷	۳۸	۳۹	۴۰	۴۱	۴۲	۴۳	۴۴	۴۵	۴۶	۴۷	۴۸
کد باینری	۰۱۱۰۰	۰۱۱۰۱	۰۱۱۱۰	۰۱۱۱۱	۱۰۰۰۰	۱۰۰۰۱	۱۰۰۱۰	۱۰۰۱۱	۱۰۱۰۰	۱۰۱۰۱	۱۰۱۱۰	۱۰۱۱۱	۱۱۰۰۰	۱۱۰۰۱	۱۱۰۱۰	۱۱۰۱۱	۱۱۱۰۰	۱۱۱۰۱	۱۱۱۱۰	۱۱۱۱۱	۱۰۰۰۰۰

همان‌طور که از جدول (۴-۳) مشاهده می‌گردد با استفاده از مدل یونیگرام تعداد بیت‌های موردنیاز برای رمز نمودن کاراکترهای پرتکرار کاهش می‌یابد. البته جهت تفکیک نمودن کاراکترهای خاص از سایر کاراکترها به یک بیت کنترلی در ابتدای هر کاراکتر نیاز می‌باشد که بسته به مقدار آن مشخص می‌کند که آیا کاراکتر نوع خاص است یا خیر. با فرض اینکه بیت کنترلی یک بیانگر کاراکتر خاص و بیت صفر بیانگر سایر کاراکترها باشد آنگاه کد باینری کاراکترها به‌صورت جدول (۴-۴) خواهد بود. بنابراین در حضور بیت کنترلی، هر کاراکتر خاص به ۴ بیت و سایر کاراکترها به ۷ بیت نیاز دارند.

جدول۴. ۴ کاراکترهای فارسی همراه با کد پیشنهادی آن‌ها با روش مدل زبانی یونیگرام همراه با بیت کنترلی(حروف الفبای فارسی، کاراکترهای انفصال و ارقام)

کاراکتر	ا	د	ر	ن	و	ه	ی	فاصله
شماره کاراکتر	۰	۱	۲	۳	۴	۵	۶	۷
کد باینری	۱۰۰۰	۱۰۰۱	۱۰۱۰	۱۰۱۱	۱۱۰۰	۱۱۰۱	۱۱۱۰	۱۱۱۱

کاراکتر	ب	پ	ت	ث	ج	چ	ح	خ	ذ	ز	ژ	س	ش	ص	ض	ط	ظ	ع	غ	ف	ق
شماره کاراکتر	۰	۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰	۱۱	۱۲	۱۳	۱۴	۱۵	۱۶	۱۷	۱۸	۱۹	۲۰
کد باینری	۰۰۰۰۰	۰۰۰۰۱	۰۰۰۱۰	۰۰۰۱۱	۰۰۱۰۰	۰۰۱۰۱	۰۰۱۱۰	۰۰۱۱۱	۰۱۰۰۰	۰۱۰۰۱	۰۱۰۱۰	۰۱۰۱۱	۰۱۱۰۰	۰۱۱۰۱	۰۱۱۱۰	۰۱۱۱۱	۱۰۰۰۰	۱۰۰۰۱	۱۰۰۱۰	۱۰۰۱۱	۱۰۱۰۰

کاراکتر	ک	گ	ل	م	نیم‌فاصله	.	،	؛	:	؟	!	۰	۱	۲	۳	۴	۵	۶	۷	۸	۹
شماره کاراکتر	۲۱	۲۲	۲۳	۲۴	۲۵	۲۶	۲۷	۲۸	۲۹	۳۰	۳۱	۳۲	۳۳	۳۴	۳۵	۳۶	۳۷	۳۸	۳۹	۴۰	۴۱
کد باینری	۱۰۱۰۱	۱۰۱۱۰	۱۰۱۱۱	۱۱۰۰۰	۱۱۰۰۱	۱۱۰۱۰	۱۱۰۱۱	۱۱۱۰۰	۱۱۱۰۱	۱۱۱۱۰	۱۱۱۱۱	۱۰۰۰۰۰	۱۰۰۰۰۱	۱۰۰۰۱۰	۱۰۰۰۱۱	۱۰۰۱۰۰	۱۰۰۱۰۱	۱۰۰۱۱۰	۱۰۰۱۱۱	۱۰۱۰۰۰	۱۰۱۰۰۱

درصورتی‌که متن پیام فقط شامل حروف الفبا و کاراکترهای انفصال باشد و اعداد در آن حضور نداشته باشند آنگاه مدل زبانی بایگرام منجر به کاهش بیشتر تعداد بیت‌ها می‌شود در این حالت تعداد کاراکترهای غیر خاص برابر با ۳۲ کاراکتر می‌شود لذا برای کد کردن آن‌ها تنها به ۵ بیت نیاز می‌باشد که همراه با بیت کنترلی ۶ بیت نیاز دارند. این موضوع در جدول (۴-۵) نشان داده‌شده است. در این حالت کد مربوط به کاراکترهای پرتکرار همانند جدول (۴-۴) باقی خواهد ماند.

با توجه به اینکه این روش رمزنگاری روش استاندارد و معین نمی‌باشد لذا باید هر دو طرف گیرنده و فرستنده الگوریتم رمزنگاری را داشته باشند چراکه در مرحله کشف پیام، عکس عمل رمزنگاری انجام خواهد شد که رشته بیت‌ها را به کاراکترها تبدیل می‌کند.

جدول۴. ۵ کاراکترهای فارسی همراه با کد پیشنهادی آن‌ها با روش مدل زبانی یونیگرام همراه بابیت کنترلی(فقط حروف الفبای فارسی و کاراکترهای انفصال)

کاراکتر	ا	د	ر	ن	و	ه	ی	فاصله
شماره کاراکتر	۰	۱	۲	۳	۴	۵	۶	۷
کد باینری	۱۰۰۰	۱۰۰۱	۱۰۱۰	۱۰۱۱	۱۱۰۰	۱۱۰۱	۱۱۱۰	۱۱۱۱

کاراکتر	ب	پ	ت	ث	ج	چ	ح	خ	ذ	ز	ژ	س	ش	ص	ض	ط	ظ	ع	غ	ف	ق
شماره کاراکتر	۰	۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰	۱۱	۱۲	۱۳	۱۴	۱۵	۱۶	۱۷	۱۸	۱۹	۲۰
کد باینری	۰۰۰۰۰	۰۰۰۰۱	۰۰۰۱۰	۰۰۰۱۱	۰۰۱۰۰	۰۰۱۰۱	۰۰۱۱۰	۰۰۱۱۱	۰۱۰۰۰	۰۱۰۰۱	۰۱۰۱۰	۰۱۰۱۱	۰۱۱۰۰	۰۱۱۰۱	۰۱۱۱۰	۰۱۱۱۱	۱۰۰۰۰	۱۰۰۰۱	۱۰۰۱۰	۱۰۰۱۱	۱۰۱۰۰

کاراکتر	ک	گ	ل	م	نیم‌فاصله	.	،	؛	:	؟	!
شماره کاراکتر	۲۱	۲۲	۲۳	۲۴	۲۵	۲۶	۲۷	۲۸	۲۹	۳۰	۳۱
کد باینری	۱۰۱۰۱	۱۰۱۱۰	۱۰۱۱۱	۱۱۰۰۰	۱۱۰۰۱	۱۱۰۱۰	۱۱۰۱۱	۱۱۱۰۰	۱۱۱۰۱	۱۱۱۱۰	۱۱۱۱۱

البته اینکه در جداول بالا به ترتیب حروف الفبا رمزنگاری انجام‌شده است یک رویکرد ثابت نیست بدین معنی که می‌توان از توالی‌های مختلف برای تخصیص مقدار به کاراکترها استفاده نمود مثلاً اینکه ازانتهای حروف الفبا شروع به تخصیص مقدار نمود و یا اینکه کاراکترهای انفصال در ابتدای توالی باشند و بعد حروف و بعد ارقام بیایند.

۴-۲-۱ کلید رمز

علاوه بر این موارد می‌توان از یک کلید رمز نیز استفاده نمود مثلاً کاراکترها به سه دسته حروف الفبا، کاراکترهای انفصال و ارقام تقسیم نمود. این سه دسته با ۶ حالت می‌توانند در توالی هم قرار بگیرند. این حالت‌ها را می‌توان به‌صورت زیر بیان نمود:

حالت صفر: حروف الفبا-کاراکترهای انفصال-ارقام (۰۰۰)

حالت یک: حروف الفبا- ارقام- کاراکترهای انفصال (۰۰۱)

حالت دو: کاراکترهای انفصال- حروف الفبا- ارقام (۰۱۰)

حالت سه: کاراکترهای انفصال- ارقام- حروف الفبا (۰۱۱)

حالت چهار: ارقام- حروف الفبا- کاراکترهای انفصال (۱۰۰)

حالت پنج: ارقام- کاراکترهای انفصال- حروف الفبا (۱۰۱)

حالت شش: حروف الفبا- کاراکترهای انفصال (۱۱۰)

حالت هفت: کاراکترهای انفصال- حروف الفبا (۱۱۱)

و اعداد صفرتا هفت را به این حالت‌ها اختصاص داد. سپس در مرحله نهان‌نگاری که در ابتدا طول رشته بیت‌ها ذخیره می‌شود بعدازآن سه بیت برای تعیین اینکه کدام حالت در رمزنگاری کاراکترها استفاده‌شده است اختصاص داد با این کار در مرحله کشف پیام تبدیل رشته بیت‌ها به کاراکترها

درست انجام می‌شود. در جدول (۴-۴) حالت صفر نشان داده‌شده است. در حالت‌های ششم و هفتم ارقام حضور ندارند لذا برای کد کردن کاراکترها باید از جدول (۴-۵) استفاده نمود.

حالت‌های مختلف برای کلید رمز می‌توان استفاده نمود مثلاً علاوه بر ۸ حالت بالا، می‌توان برای خود دسته‌ها نیز کلید تعریف نمود بدین‌صورت مثلاً در دسته حروف الفبا اینکه ابتدا کدام حرف قرار بگیرد را می‌توان در کلید تعریف نمود. این کلید باید بعد از کلید حالت به‌صورت یک‌رشته بیت ۵ تایی (اعداد صفرتا ۳۱) بیاید که درواقع عدد معادل بیانگر این است که کدام حرف الفبا در ابتدای دسته حروف الفبا قرارگرفته است. مثلاً اگر کلید عدد ۴ (۰۰۱۰۰) باشد آنگاه ترتیب حروف الفبا به‌صورت توالی زیر خواهد بود:

ت ث ج چ ح خ د ذ ر ز ژ س ش ص ض ط ظ ع غ ف ق ک گ ل م ن و ه ی ا ب پ

یا برای دسته ارقام یک‌رشته بیت ۴ تایی که معرف رقم اول دسته می‌باشد در نظر گرفت. مثلاً اگر این عدد ۳ (۰۰۱۱) باشد آنگاه توالی ارقام به‌صورت زیر خواهد بود:

"۲ ۱ ۰ ۹ ۸ ۷ ۶ ۵ ۴ ۳"

در حالت‌های شش و هفت که ارقام حضور ندارند عدد این کلید مهم نیست و می‌توان به‌صورت پیش‌فرض آن را صفر (۰۰۰۰) در نظر گرفت.

همچنین برای دسته کاراکترهای انفصال (۸ کاراکتر انفصال)، یک کلید سه بیت تعریف نمود که بازهم عدد موردنظر ترتیب توالی را مشخص کند. به‌عنوان قرارداد توالی تعریف‌شده در جدول را به‌عنوان ترتیب صفر تعریف می‌کنیم. لذا اگر کلید ۳ بیتی مقدار ۴ (۱۰۰)داشته باشد آنگاه توالی کاراکترهای انفصال به‌صورت زیر خواهد بود:

؛	:	؟	!	٠	فاصله	نیم‌فاصله	.	،

بنابراین به‌صورت کلی یک کلید رمز ۴ قسمتی تعریف خواهد شد که قسمت اول آن‌یک رشته بیت ۳ تایی می‌باشد که معرف شماره حالت توالی دسته می‌باشد و قسمت‌های دوم، سوم و چهارم، بسته نوع حالت، کلیدهای معرف توالی کاراکترها در هر دسته می‌باشند. این کلید کلاً ۱۵ بیت طول دارد(شماره حالت ۳بیت، توالی حروف الفبا ۵ بیت، توالی کاراکترهای انفصال ۳ بیت و ارقام ۴ بیت).

در اینجا دو نمونه کلید نوعی بررسی می‌شود:

کلید: حالت۵- توالی ۳-توالی ۴- توالی ۱۶

این کلید بیان می‌کند که ترتیب توالی دسته کاراکترها به‌صورت ارقام- کاراکترهای انفصال- حروف الفبا می‌باشد، ترتیب کاراکترهای ارقام از "۲" شروع می‌شود، ترتیب کاراکترهای انفصال از "،" و ترتیب حروف الفبا از حرف "ش" شروع می‌شود. در جدول (۴-۶) رمزنگاری کاراکترها با این کلید نشان داده‌شده است.

جدول۴. ۶ رمزنگاری کاراکترها با کلید نوعی: حالت۵- توالی ۳-توالی ۴- توالی ۱۶

کاراکتر	۲	۳	۴	۵	۶	۷	۸	۹	۰	۱	،	؛	:	؟	!	فاصله	نیم‌فاصله	.	ش	ص	ض	ط	ظ	ع	غ
شماره کاراکتر	۰	۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰	۱۱	۱۲	۱۳	۱۴	۱۵	۱۶	۱۷	۱۸	۱۹	۲۰	۲۱	۲۲	۲۳	۲۴
کد باینری	۰۰۰۰	۰۰۰۱	۰۰۱۰	۰۰۱۱	۰۱۰۰	۰۱۰۱	۰۱۱۰	۰۱۱۱	۱۰۰۰	۱۰۰۱	۱۰۱۰	۱۰۱۱	۱۱۰۰	۱۱۰۱	۱۱۱۰	۱۱۱۱	۰۰۰	۰۰۱	۰۱۰	۰۱۱	۱۰۰	۱۰۱	۱۱۰	۱۱۱	۰۰۰۰	۰۰۰۱

کاراکتر	ف	ق	ک	گ	ل	م	ن	و	ه	ی	ا	ب	پ	ت	ث	ج	چ	ح	خ	د	ذ	ر	ز	ژ	س
شماره کاراکتر	۲۵	۲۶	۲۷	۲۸	۲۹	۳۰	۳۱	۳۲	۳۳	۳۴	۳۵	۳۶	۳۷	۳۸	۳۹	۴۰	۴۱	۴۲	۴۳	۴۴	۴۵	۴۶	۴۷	۴۸	۴۹
کد باینری	۰۰۱۱۰۰	۰۰۱۱۰۱	۰۰۱۱۱۰	۰۰۱۱۱۱	۰۰۱۱۰۰	۰۰۱۱۰۱	۰۰۱۱۱۰	۰۰۱۱۱۱	۰۰۱۱۰۰	۰۰۱۱۰۱	۰۰۱۱۱۰	۰۰۱۱۱۱	۰۰۱۱۰۰	۰۰۱۱۰۱	۰۰۱۱۱۰	۰۰۱۱۱۱	۰۰۱۱۰۰	۰۰۱۱۰۱	۰۰۱۱۱۰	۰۰۱۱۱۱	۰۰۱۱۰۰	۰۰۱۱۰۱	۰۰۱۱۱۰	۰۰۱۱۱۱

بعد از ۱۶ بیت ابتدایی که درواقع بیانگر طول رشته بیت‌های پیام می‌باشد کلید بالا به‌صورت رشته بیت جدول (۴-۷) خواهد بود:

جدول۴. ۷ رشته بیت کلید نوعی: حالت۵- توالی ۳-توالی ۴- توالی ۱۶

۱۰۰۰۰	۱۰۰	۰۰۱۱	۱۰۱	۱۶ بیت اول (طول رشته بیت‌های پیام)

جدول۴. ۸ رمزنگاری کاراکترها با کلید نوعی: حالت۶- توالی ۳-توالی ۱- توالی ۰

کاراکتر	د	ر	ن	و	ه	ی	ا	فاصله
شماره کاراکتر	۰	۱	۲	۳	۴	۵	۶	۷
کد باینری	۱۰۰۰	۱۰۰۱	۱۰۱۰	۱۰۱۱	۱۱۰۰	۱۱۰۱	۱۱۱۰	۱۱۱۱

کاراکتر	پ	ت	ث	ج	چ	ح	خ	ذ	ز	ژ	س	ش	ص	ض	ط	ظ	ع	غ	ف	ق	ک
شماره کاراکتر	۰	ـ	ـ	ـ	ـ	ـ	ـ	ـ	ـ	ـ	ـ	ـ	ـ	ـ	ـ	ـ	ـ	ـ	ـ	ـ	ـ
کد باینری	۰۰۰۰۰۰	۰۰۰۰۰۱	۰۰۰۰۱۰	۰۰۰۰۱۱	۰۰۰۱۰۰	۰۰۰۱۰۱	۰۰۰۱۱۰	۰۰۰۱۱۱	۰۰۱۰۰۰	۰۰۱۰۰۱	۰۰۱۰۱۰	۰۰۱۰۱۱	۰۰۱۱۰۰	۰۰۱۱۰۱	۰۰۱۱۱۰	۰۰۱۱۱۱	۰۰۱۱۰۰	۰۰۱۱۰۱	۰۰۱۱۱۰	۰۰۱۱۱۱	۰۰۱۱۰۰

کاراکتر	گ	ل	م	ب	نیم‌فاصله	.	،	؛	:	؟	!
شماره کاراکتر	۲۱	۲۲	۲۳	۲۴	۲۵	۲۶	۲۷	۲۸	۲۹	۳۰	۳۱
کد باینری	۱۰۰۰۰۰	۱۰۰۰۱۰	۱۰۰۰۱۱	۱۰۰۱۰۰	۱۰۰۱۰۱	۱۰۰۱۱۰	۱۰۰۱۱۱	۱۰۱۰۰۰	۱۰۱۰۰۱	۱۰۱۰۱۰	۱۰۱۰۱۱

جدول۴. ۹رشته بیت کلید نوعی: حالت۶- توالی ۳-توالی ۱- توالی صفر

۰۰۰۰	۰۰۱۰۱	۰۱۱	۰۱۰	۱۶ بیت اول (طول رشته بیت‌های پیام)

حضور این کلید باعث افزایش امنیت رمزنگاری خواهد شد. البته نحوه تعریف الگوریتم کلید باید هم برای طرف فرستنده و هم طرف گیرنده معلوم باشد.

برای نهان‌نگاری پیام در متن از قبل باید رشته بیت‌های پیام آماده باشند که این رشته بیت‌ها شامل ۱۶ بیت اول برای طول رشته بیت‌ها، ۱۵ بیت بعدی برای کلید رمزنگاری و در ادامه بیت‌هایی می‌آیند که متناظر با کاراکترهای پیام می‌باشند. این مجموعه بیت در یک دنباله (صف) قرار داده می‌شوند. کاراکترهای پیام پشت سر هم به همان ترتیب موجود در پیام رمزنگاری می‌شوند. در اینجا رشته بیت‌ها برای یک پیام نوعی با استفاده از روش پیشنهادی بدست آورده می‌شود..

پیام: "جلسه فردا تشکیل می‌گردد."

با شروع از ابتدای پیام کاراکترها تبدیل به کد می‌شوند. برای تبدیل کاراکترها به کد باید کلید حالت رمزنگاری تعیین شود.

کلید: حالت شش- توالی ۳- توالی صفر (۱۱۰۰۰۰۱۱۰۰۰۰۰۰۰)

کد کاراکترهای پیام بالا به‌صورت جدول (۴-۱۰) می‌باشد:

جدول۴. ۱۰ کد کاراکترهای پیام: "جلسه فردا تشکیل می‌گردد"

کاراکتر	ج	ل	س	ه	فاصله	ف	ر	د
کد	۰۰۰۰۱۱	۰۱۰۱۱۰	۰۰۱۰۱۰	۱۱۰۰	۱۱۱۱	۰۱۰۰۱۰	۱۰۰۱	۱۰۰۰
کاراکتر	ا	فاصله	ت	ش	ک	ی	ل	فاصله
کد	۱۱۱۰	۱۱۱۱	۰۰۰۰۰۱	۰۰۱۰۱۱	۰۱۰۱۰۰	۱۱۰۱	۰۱۰۱۱۰	۱۱۱۱
کاراکتر	م	ی	نیم‌فاصله	گ	ر	د	د	.
کد	۰۱۰۱۱۱	۱۱۰۱	۰۱۱۰۰۱	۰۱۰۱۰۱	۱۰۰۱	۱۰۰۰	۱۰۰۰	۰۱۱۰۱۰

همان‌طور که در جدول (۴-۱۰) نشان داده‌شده است طول رشته بیت‌های پیام موردنظر برابر ۱۲۰ بیت می‌باشد. لذا در ۱۶ بیت اول رشته بیت‌ها باید معادل باینری عدد ۱۲۰ (۰۰۰۰۰۰۰۰۱۱۱۱۰۰۰) قرار داده شود و بعدازآن معادل باینری کلید (۱۱۰۰۰۰۱۱۰۰۰۰۰۰۰) آورده شود و سپس بیت‌های معادل کاراکترها در دنباله آورده خواهند شد. لذا رشته بیت لازم برای نهان‌نگاری پیام بالا به‌صورت زیر صف زیر خواهد بود. صف دارای ۱۵۱ بیت می‌باشد.

“

۰۰۰۰۰۰۰۰۱۱۱۱۰۰۰۱۱۰۰۰۰۱۱۰۰۰۰۰۰۰۰۰۰۱۱۰۱۰۱۱۰۰۰۱۰۱۰۱۱۰۰۱۱۱۱۰۱۰۰۱۰۱

۰۰۱۱۰۰

۰۱۱۱۰۱۱۱۱۰۰۰۰۰۱۰۰۱۰۱۱۱۰۱۰۰۱۱۰۱۰۱۰۱۱۰۱۱۱۱۰۱۰۱۱۱۱۱۰۱۰۱۱۰۰۱۰۱۰۱۰۱۱۰

”۰۱۱۰۰۰۱۰۰۰۰۱۱۰۱۰

علاوه بر کلید تعریف‌شده برای رمزنگاری کاراکترها، جهت افزایش امنیت رمزنگاری می‌توان به جای اینکه کاراکترهای پیام به‌صورت پشت سر هم در رشته بیت قرار بگیرند آن‌ها را به‌صورت تصادفی با استفاده از یک تابع تصادفی با عنوان کلید در کنار هم قرارداد و آن را در متن پنهان نمود.

مثلاً در همان پیام بالا که مجموعاً ۲۴ کاراکتر دارد هر کاراکتر یک شماره از ۱ تا ۲۴ بگیرد و در موقعیت اختصاص یافته قرار بگیرد. مثلاً فرض شود که تابع تصادفی زیر برای فاصله ۱ تا ۲۴ تعریف‌شده باشد.

[20 4 8 3 9 11 15 18 22 24 16 12 19 6 1 17 2 14 10 23 7 5 13 21]

در آن صورت ترتیب قرار گرفتن کاراکترها در رمزنگاری به‌صورت زیر خواهد بود:

ی	ل	ت	ا	س	د	ه	گ	کاراکتر
۱۱۰۱	۰۱۰۱۱۰	۰۰۰۰۰۱	۱۱۱۰	۰۰۱۰۱۰	۱۰۰۰	۱۱۰۰	۰۱۰۱۰۱	کد
م	ج	ف	نیم‌فاصله	ش	فاصله	.	د	کاراکتر
۰۱۰۱۱۱	۰۰۰۰۱۱	۰۱۰۰۱۰	۰۱۱۰۰۱	۰۰۱۰۱۱	۱۱۱۱	۰۱۱۰۱۰	۱۰۰۰	کد
ر	ک	فاصله	ر	د	فاصله	ی	ل	کاراکتر
۱۰۰۱	۰۱۰۱۰۰	۱۱۱۱	۱۰۰۱	۱۰۰۰	۱۱۱۱	۱۱۰۱	۰۱۰۱۱۰	کد

در این حالت پیام به‌صورت تصادفی در متن پنهان می‌شود. لذا در زمان کشف پیام، کاراکترها به‌صورت تصادفی کشف می‌شوند پس الگوریتم تابع تصادفی (کلید) باید برای طرف گیرنده هم معلوم باشد.

در این بخش پیام با استفاده مدل زبانی یونیگرام و به کمک کلید رمز تعریف‌شده به کدهای باینری با حداقل بیت موردنیاز و با امنیت بالای رمزنگاری تبدیل شد. در بخش بعدی روش نهان کردن این بیت‌ها در متن فارسی بیان خواهد شد.

۳-۴ نهان‌نگاری

در این بخش نهان‌نگاری پیام در متن فارسی به کمک ویژگی‌های کاراکتر نیم‌فاصله و همچنین ترکیب ویژگی‌های کاراکترهای نیم‌فاصله و فاصله انجام می‌شود. درواقع در این مرحله پیام رمزنگاری‌شده که به‌صورت رشته بیت‌های باینری در آمده است در یک متن فارسی پنهان می‌شود. در کد باینری هر بیت دو حالت صفر و یک دارد لذا برای ذخیره هر بیت در متن می‌توان از تغییر یا

عدم‌تغییر یک ویژگی از کاراکترهای موجود در متن استفاده نمود بدین‌صورت که به‌صورت قراردادی یکی از دو تغییر یا عدم‌تغییر ویژگی به‌عنوان بیت یک و دیگری به‌عنوان بیت صفر لحاظ شود. در این پروژه برای نهان کردن بیت‌ها، از ویژگی‌های کاراکتر نیم‌فاصله استفاده می‌شود که دلیل انتخاب این کاراکتر، هم مشخص نبودن تغییر شکل ظاهری این کاراکتر و هم فراوانی تکرار آن در متون فارسی می‌باشد.

۴-۳-۱ کاراکتر نیم‌فاصله

در زبان فارسی تعدادی از کلمات به‌صورت دوقسمتی می‌باشند، مانند "می‌باشند"، که برای نوشتن آن‌ها به‌منظور جدا نشدن این دو کلمه، از کاراکتر نیم‌فاصله استفاده می‌شود. این کاراکتر اغلب در فعل‌ها و یا اسامی جمع دیده می‌شود. تکرار این کاراکتر در زبان فارسی کم می‌باشد اما ازآنجایی‌که تعداد و تغییر ویژگی این کاراکتر تغییری در شکل ظاهری متن ایجاد نمی‌کند لذا کاراکتر مناسبی برای عمل نهان‌نگاری می‌باشد. به همین دلیل در این پایان‌نامه با ایجاد کاراکترهای اضافی نیم‌فاصله، تعداد آن‌ها در متن به شدت اضافه‌شده به‌طوری‌که ظرفیت نهان‌نگاری با استفاده از کاراکتر نیم‌فاصله بسیار بالا رفته است.

۴-۳-۱-۱ ایجاد کاراکتر نیم‌فاصله اضافی در متن

در نوشتار زبان فارسی تعدادی از کاراکترها به کاراکتر قبلی می‌چسبند و تعدادی دیگر به‌صورت جدا نوشته می‌شوند مثلاً در کلمه "می‌خورد"، حرف "ی" به حرف "م" چسبیده درحالی‌که حرف "خ" جدا از حرف "ی" نوشته‌شده است و حرف "و" به حرف "خ" چسبیده و حرف "ر" جدا از حرف "و" نوشته‌شده است و حرف "د" نیز جدا از حرف "ر" می‌باشد. لذا در فاصله حروفی که به هم نمی‌چسبند می‌توان به تعداد دلخواه کاراکتر نیم‌فاصله قرارداد بدون آنکه تغییری در شکل ظاهری متن ایجاد شود. مثلاً در همین کلمه "می‌خورد"، به‌صورت معمول یک کاراکتر نیم‌فاصله بین بخش

"می" و بخش "خورد" وجود دارد اما می‌توان بین حروف غیرچسبان نیز کاراکتر نیم‌فاصله قرارداد. لذا کلمه "می‌خورد" به‌صورت زیر درمی‌آید :

"م ی نیم‌فاصله خ و نیم‌فاصله ر نیم‌فاصله د"

بنابراین در این کلمه علاوه بر یک کاراکتر معمول نیم‌فاصله، می‌توان دو کاراکتر نیم‌فاصله دیگر ایجاد نمود. در جدول (۴-۱۱)تعداد کاراکترهای قابل ایجاد برای چند کلمه نمونه نشان داده‌شده است.

جدول۴. ۱۱ امکان ایجاد کاراکتر نیم‌فاصله برای چند کلمه نوعی

کلمه	ایران	آسمان	مازندران	ارقام
تعداد کاراکتر نیم‌فاصله	۳	۲	۵	۳

۴-۳-۲ نهان‌نگاری با استفاده از ویژگی‌های کاراکتر نیم‌فاصله

قلمی که برای نوشتار استفاده می‌شود دارای ویژگی‌های مختلفی می‌باشد این ویژگی‌ها شامل نوع، اندازه، ضخامت، رنگ و خمیدگی قلم می‌باشند. لذا کاراکترهای نیم‌فاصله می‌توانند انواع مختلفی از ویژگی‌ها را داشته باشند. در یک متن معمولاً از یک فونت یکسان برای همه کاراکترها استفاده می‌شود و یا حداقل در هر پاراگراف یا جمله و درنتیجه در هر کلمه از فونت یکسانی برای نوشتن کاراکترها استفاده می‌شود. حال برای پنهان نمودن بیت‌های باینری از تغییر ویژگی‌های فونت کاراکتر نیم‌فاصله نسبت به کاراکتر ماقبل آن در هر کلمه استفاده می‌شود.

در یک متن می‌توان از تغییر یکی یا چند ویژگی کاراکتر نیم‌فاصله برای نهان نمودن بیت‌ها استفاده نمود. اینکه تغییر یا عدم‌تغییر ویژگی کاراکتر به‌عنوان بیت صفر یا یک تعیین شود وابسته به قراردادی است که بین فرستنده و گیرنده وجود دارد. مثلاً تغییر ویژگی به‌عنوان بیت یک و عدم‌تغییر ویژگی به‌عنوان بیت صفر تعیین شود.

برای نهان‌نگاری از اول متن شروع به نهان‌نگاری می‌شود و با رسیدن به هر کاراکتر نیم‌فاصله، با توجه به بیت ابتدای صف (یک یا صفر)، ویژگی آن یا تغییر داده می‌شود و یا ثابت می‌ماند. همچنین جهت افزایش مقاومت نهان‌نگاری، چندین بار نهان‌نگاری در متن انجام می‌شود بدین‌صورت که بعد از اتمام هر بار نهان‌نگاری، تکرار بعدی شروع می‌شود. بنابراین در این روش مقاومت نهان‌نگاری در مقابل وجود خطاهای احتمالی افزایش می‌یابد. در مرحله کشف پیام، با مقایسه پیام‌های کشف‌شده، احتمال دست‌یابی به پیام صحیح افزایش می‌یابد.

در جدول (۴-۱۲) نهان‌نگاری پیام عنوان‌شده در بخش قبلی ("فردا جلسه برگزار می‌گردد") در متن فارسی با استفاده از تغییر ویژگی‌فونت کاراکتر نیم‌فاصله نشان داده‌شده است.

جدول۴. ۱۲ نهان‌نگاری متن فارسی با تغییر ویژگی فونت کاراکتر نیم‌فاصله نسبت به کاراکتر قبلی

متن	متن نهان‌نگاری شده(فونت)
تولید این نرم‌افزار به‌منظور اتوماسیون نمودن امورات مربوط به نظارت در طول مراحل بازدید، تهیه گزارش تخلف و گزارش‌های دوره‌ای می‌باشد. به‌طوری که این نرم افزار می‌تواند با داشتن یک پایگاه داده مناسب اطلاعات مربوط به دوره‌های مختلف را در خود ذخیره نموده و در زمان نیاز این اطلاعات به آسانی قابل بازیابی و استفاده باشند. همچنین امکان دسته‌بندی و مرتب نمودن داده‌ها برحسب نیاز برای اهداف متفاوت قابل حصول می‌باشد. از بارزترین ویژگی‌های این نرم‌افزار می‌توان به امکان ذخیره اطلاعات مربوط به ساخت وساز در پایگاه داده، دسترسی سریع به آمار و اطلاعات مربوط به ساخت و ساز، میزان رعایت یا عدم رعایت مقررات ملی ساختمان در زیربخش‌های مختلف در سطح استان و به تفکیک شهرستان، ارائه نمودارهای گوناگون از وضعیت مواد مختلف مقررات ملی، امکان وارد نمودن اطلاعات در حین	تولید این نرم‌افزار به‌منظور اتوماسیون نمودن امورات مربوط به نظارت در طول مراحل بازدید، تهیه گزارش تخلف و گزارش‌های دوره‌ای می‌باشد. به‌طوری که این نرم افزار می‌تواند با داشتن یک پایگاه داده مناسب اطلاعات مربوط به دوره‌های مختلف را در خود ذخیره نموده و در زمان نیاز این اطلاعات به آسانی قابل بازیابی و استفاده باشند. همچنین امکان دسته‌بندی و مرتب نمودن داده‌ها برحسب نیاز برای اهداف متفاوت قابل حصول می‌باشد. از بارزترین ویژگی‌های این نرم‌افزار می‌توان به امکان ذخیره اطلاعات مربوط به سا از در پایگاه داده، دسترسی سریع به آم عات مربوط به ساخت و ساز، میزان عدم رعایت مقررات ملی ساختمان بخش‌های مختلف در سطح استان و به تفکیک شهرستان، ارائه نمودارهای گوناگون از وضعیت مواد مختلف مقررات ملی، امکان وارد نمودن اطلاعات در حین

اتمام نهان‌نگاری

بازدیدهای میدانی و همچنین امکان ارسال گزارش تخلف به اتوماسیون اداری به‌صورت آنلاین، امکان ثبت موقعیت مکانی پروژه‌های مورد بازدید اشاره نمود.	بازدیدهای میدانی و همچنین امکان ارسال گزارش تخلف به اتوماسیون اداری به‌صورت آنلاین، امکان ثبت موقعیت مکانی پروژه‌های مورد بازدید اشاره نمود.

همان‌طور که مشاهده می‌گردد نهان‌نگاری با استفاده کاراکتر نیم‌فاصله هیچ تغییر ظاهری در شکل متن ایجاد نمی‌کند که در مقایسه با روش‌هایی که بر روی تغییر فاصله خطوط، جابجایی افقی پاراگراف‌ها، تغییر شیب حروف منحنی مزیت مهمی به‌حساب می‌آید. در متن نهان‌نگاری شده علاوه بر پیام، طول رشته بیت‌ها و کلید رمزنگاری نیز پنهان‌شده است.

جهت افزایش بیشتر امنیت نهان‌نگاری، علاوه بر کلید موجود در رمزنگاری پیام، می‌توان در مرحله نهان‌نگاری الگوریتم پیچیده‌تری تعریف نمود. یک الگوریتم می‌تواند بدین‌صورت باشد که بیت صفر و یک را برای تغییر یا عدم‌تغییر ویژگی در قسمتی از متن جابجا نمود. مثلاً برای ۵۰ درصد رشته بیت‌ها، تغییر ویژگی به معنای بیت یک و برای ۵۰ درصد دیگر، تغییر ویژگی به مفهوم بیت صفر باشد. و یا مثلاً می‌توان قرارداد نمود که به‌صورت یک‌درمیان مفاهیم بیت صفر و یک جابجا شود. و یا هر نوع الگوریتم دیگری که می‌تواند به امنیت بیشتر نهان‌نگاری کمک نماید.

از طرف دیگر با توجه به اینکه هکرها متون مختلف را تجزیه‌وتحلیل می‌نمایند لذا وجود تغییرات مشخص در متن، وجود پیام و همچنین کشف آن را ساده‌تر می‌نماید. بنابراین به‌منظور پیچیده نمودن الگوریتم نهان‌نگاری می‌توان تغییرات متعددی در متن ایجاد نمود و در الگوریتم یک کلید تعریف نمود که مشخص نماید که کدام تغییر باید در مرحله کشف پیام تأثیر داده شود. این کلید به‌صورت زیر تعریف می‌شود:

حالت صفر: تغییر ویژگی فونت (۰۰۰)

حالت یک: تغییر ویژگی سایز (۰۰۱)

حالت دو: تغییر ویژگی ضخامت (۰۱۰)

حالت سه: تغییر ویژگی رنگ (۰۱۱)

حالت چهار: تغییر ویژگی خمیدگی (۱۰۰)

حالت پنج: تغییر همه ویژگی‌ها (۱۰۱)

این کلید می‌تواند همراه کلید رمزنگاری و همراه کلید الگوریتم نهان‌نگاری (مثلاً اینکه به‌صورت یک‌درمیان مفهوم بیت‌های صفر و یک عوض شود) در رشته بیت‌های پیام آورده شود.

۴-۳-۳ نهان‌نگاری به کمک کاراکترهای نیم‌فاصله و فاصله

جهت بالا بردن ظرفیت نهان‌نگاری، می‌توان از ظرفیت هم‌زمان کاراکتر نیم‌فاصله و فاصله استفاده نمود که باعث می‌شود جهت رمزنگاری به متون کوتاه‌تری نیاز باشد. این مسئله می‌تواند به تکرارهای زیاد نهان‌نگاری در متون کوتاه نیز کمک کند که خود باعث افزایش مقاومت نهان‌نگاری می‌شود. نهان‌نگاری به کمک دو کاراکتر نیم‌فاصله و فاصله در جدول (۴-۱۳) نشان داده‌شده است. همان‌طور که مشاهده می‌گردد نهان‌نگاری با استفاده از دو کاراکتر نیم‌فاصله و فاصله به‌صورت هم‌زمان، به متن کوتاهتری نیاز دارد لذا ظرفیت نهان‌نگاری متن افزایش می‌یابد. بنابراین در این روش امکان تکرار چند باره پیام در متون کوتاه نیز وجود دارد.

جدول۴. ۱۳ نهان‌نگاری به کمک کاراکترهای نیم‌فاصله و فاصله

متن	متن نهان‌نگاری شده(فونت)
تولید این نرم‌افزار به‌منظور اتوماسیون نمودن امورات مربوط به نظارت در طول مراحل بازدید، تهیه گزارش تخلف و گزارش‌های دوره‌ای می‌باشد. به‌طوری که این نرم افزار می‌تواند با داشتن یک پایگاه داده مناسب اطلاعات مربوط به دوره‌های مختلف را در خود ذخیره نموده و در زمان نیاز این اطلاعات به آسانی قابل بازیابی و استفاده باشند. همچنین امکان دسته‌بندی و مرتب نمودن داده‌ها برحسب نیاز برای اهداف متفاوت قابل حصول می‌باشد. از بارزترین ویژگی‌های این نرم‌افزار می‌توان به امکان ذخیره اطلاعات مربوط به ساخت وساز در پایگاه داده، دسترسی سریع به آمار و اطلاعات مربوط به ساخت و ساز، میزان رعایت یا عدم رعایت مقررات ملی ساختمان در زیربخش‌های مختلف در سطح استان و به تفکیک شهرستان، ارائه نمودارهای گوناگون از وضعیت مواد مختلف مقررات ملی، امکان وارد نمودن اطلاعات در حین بازدیدهای میدانی و همچنین امکان ارسال گزارش تخلف به اتوماسیون اداری به‌صورت آنلاین، امکان ثبت موقعیت مکانی پروژه‌های مورد بازدید اشاره نمود.	تولید این نرم‌افزار به‌منظور اتوماسیون نمودن امورات مربوط به نظارت در طول مراحل بازدید، تهیه گزارش تخلف و گزارش‌های دوره‌ای می‌باشد. به‌طوری که این نرم افزار می‌تواند با داشتن یک پایگاه داده مناسب اطلاعات مربوط به دوره‌های مختلف را در خود ذخیره نموده و در زمان نیاز این اطلاعات به آسانی قابل بازیابی و ا شند. همچنین امکان دسته‌بندی و م دن داده‌ها برحسب نیاز برای اهداف م ابل حصول می‌باشد. از بارزترین و ی این نرم‌افزار می‌توان به امکان ذخیره اطلاعات مربوط به ساخت وساز در پایگاه داده، دسترسی سریع به آمار و اطلاعات مربوط به ساخت و ساز، میزان رعایت یا عدم رعایت مقررات ملی ساختمان در زیربخش‌های مختلف در سطح استان و به تفکیک شهرستان، ارائه نمودارهای گوناگون از وضعیت مواد مختلف مقررات ملی، امکان وارد نمودن اطلاعات در حین بازدیدهای میدانی و همچنین امکان ارسال گزارش تخلف به اتوماسیون اداری به‌صورت آنلاین، امکان ثبت موقعیت مکانی پروژه‌های مورد بازدید اشاره نمود.

اتمام نهان‌نگاری

۴-۴ کشف پیام

پیام نهان شده در طرف گیرنده باید استخراج شود به‌منظور استخراج پیام، هم الگوریتم نهان‌نگاری و هم الگوریتم رمزنگاری و هم کلیدها باید برای طرف گیرنده معلوم باشد. در مرحله اول با توجه به الگوریتم و کلیدهای نهان‌نگاری، بیت‌های صفر و یک موجود در متن استخراج می‌شوند و سپس با توجه به الگوریتم و کلیدهای رمزنگاری رشته بیت‌های صفر و یک به کاراکترهای پیام تبدیل می‌شوند. در این قسمت یک متن نهان‌نگاری شده نشان داده‌شده است که برای این متن پیام نهان شده را استخراج می‌کنیم.

تحلیل نقشه‌های هواشناسی نشان می‌دهد با ورود یک سامانه بارشی به کشور، دوشنبه افزون بر غرب و شمال غرب بتدریج در برخی منطقه‌های جنوب، مرکز و استان‌های واقع در دامنه های جنوبی البرز شاهد بارش برف و باران، وزش باد و در بخش های جنوبی رعد و برق خواهیم بود. سازمان هواشناسی کشور اعلام کرد دوشنبه در جنوب استان های آذربایجان غربی و شرقی، کردستان، کرمانشاه، ایلام، لرستان، خوزستان، چهارمحال بختیاری و کهگیلویه و بویراحمد بارش های شدیدی رخ می دهد. با ورود سامانه بارشی از غرب و شمال غرب به کشور از یکشنبه بارش باران و برف و وزش باد در این منطقه ها آغاز شده است. همچنین سه شنبه از شدت بارش در غرب، جنوب و مرکز کشور کاسته و بارش به استان های شرقی و شمال شرقی کشور محدود می‌شود.

برای پیام بالا ابتدا با توجه به الگوریتم نهان‌نگاری (مثلاً فرض شود که الگوریتم تغییر نوع فونت کاراکتر به‌عنوان بیت یک و عدم‌تغییر به‌عنوان بیت صفر) رشته بیت‌های صفر و یک حاصل می‌شوند. ۱۶ بیت اول به معنای طول رشته بیت و ۱۵ بیت بعدی برای کلید رمزنگاری استفاده می‌شود. با توجه به ۱۶ بیت اول که معادل عدد ۳۰ می‌باشد لذا طول پیام ۳۰ بیت دارد پس بعد از ۱۵ بیت بعدی، ۳۰ بیت پیام شروع می‌شود. رشته بیت‌های صفر و یک به‌صورت زیر می‌باشد.

۰۰۰۰۰۰۰۰۰۰۱۱۱۱۰۱۱۰۰۰۰۱۱۰۰۱۰۰۰۰۰۰۰۰۱۱۰۱۱۱۱۰۰۱۰۱۰۱۱۰۰۰۱۰۰۰۱۱۱۰

حال بر اساس الگوریتم رمزنگاری و کلیدها رشته بیت به کاراکترها تبدیل می‌شود. با توجه به ۳ بیت بعد ۱۶ بیت اول، کلید رمزنگاری حالت شش می‌باشد همچنین ۵ بیت بعدی مشخص می‌سازد که توالی حروف ۳ می‌باشد و با توجه به ۳ بیت بعدی، توالی کاراکترها ۱ می‌باشد. ۴ بیت بعدی می‌رساند که توالی ارقام صفر می‌باشد. البته با توجه به حالت شش، ارقام در پیام وجود ندارد. بعد از ۳۱ بیت اول (مجموع ۱۶ بیت برای طول رشته و ۱۵ بیت برای حالت رمزنگاری)، ۳۰ بیت پیام را جدا می‌سازیم. سپس دسته بیت‌ها به کاراکترها تبدیل می‌شوند (از جدول (۴-۶) برای تبدیل حالت ۶ استفاده‌شده است). عکس تبدیل کدهای باینری به کاراکترها در جدول (۴-۱۴) نشان داده‌شده است.

جدول۴. ۱۴ تبدیل کدهای باینری به کاراکترهای پیام

۱۱۱۰	۰۰۱۰۰۰	۰۱۰۱۱۰	۱۰۰۱	۱۰۱۱	۰۰۰۰۰۱
ا	ز	م	ر	و	ت

لذا پیام نهان شده در متن بالا به‌صورت کلمه "تورم‌زا" می‌باشد.

۴-۵ شاخص‌های نهان‌نگاری

در این بخش شاخص‌های نهان‌نگاری بررسی‌شده و با سایر روش‌های نهان‌نگاری مقایسه خواهند شد. سه شاخص نهان‌نگاری که برای ارزیابی نهان‌نگاری تعریف‌شده‌اند شامل ظرفیت، امنیت و مقاومت نهان‌نگاری می‌باشد. در این بخش این شاخص‌ها به تفکیک بررسی می‌شوند.

۴-۵-۱ شاخص ظرفیت

شاخص ظرفیت نهان‌نگاری به‌صورت میزان قابلیت ظرفیت ذخیره اطلاعات در متن می‌باشد. با توجه به اینکه روش‌های مختلف نهان‌نگاری از ویژگی‌های متن برای ذخیره و نهان کردن اطلاعات استفاده می‌کنند لذا فرکانس تکرار ویژگی در متن می‌تواند معیار مناسبی برای ظرفیت نهان‌نگاری باشد. مثلاً درروش نهان‌نگاری به کمک جابجایی افقی یا عمودی خطوط یا پاراگراف‌ها، با توجه به

اینکه تکرار خطوط بیشتر از تکرار پاراگراف‌ها می‌باشد لذا ظرفیت آن بیشتر می‌باشد. از طرف دیگر روش‌هایی که بر روی نهان‌نگاری به کمک کلمات مانند استفاده از کلمات مترادف، تمرکز دارند از ظرفیت بیشتری نسبت به خطوط و پاراگراف‌ها برخوردار هستند. و نهایتاً روش‌هایی که از ویژگی‌های حروف و کاراکترها استفاده می‌کنند با طبع ظرفیت بیشتری دارند. در میان روش‌های نهان‌نگاری با استفاده از ویژگی‌های حروف که درگذشته بر روی آن‌ها کارشده است می‌توان به روش تغییر شیب حروف منحنی، تغییر میزان کشیدگی حروف و استفاده از کاراکتر فاصله اشاره نمود. که با توجه به تکرار بیشتر کاراکتر فاصله، روش نهان‌نگاری به کمک کاراکتر فاصله به نسبت سایر روش‌ها از ظرفیت بیشتری برخوردار است. در این میان در این پروژه با توجه به تعریف کاراکتر نیم‌فاصله مابین حروف غیرچسبان، فرکانس تکرار کاراکتر نیم‌فاصله به میزان خیلی زیادی اضافه‌شده است لذا نهان‌نگاری به کمک ویژگی‌های کاراکتر نیم‌فاصله با ظرفیت نسبتاً بالا مطرح شد. همچنین با ترکیب استفاده هم‌زمان از کاراکترهای فاصله و نیم‌فاصله، به ظرفیت بسیار بالایی برای نهان‌نگاری در این پروژه دست‌یافته شد. در جدول (۴-۱۵) ظرفیت روش‌های مختلف نهان‌نگاری به کمک فرکانس تکرار ویژگی مورداستفاده برای متون مختلف باهم مقایسه شده است. همان‌طور که ملاحظه می‌شود درروش پیشنهادی استفاده هم‌زمان از دو کاراکتر فاصله و نیم‌فاصله، ظرفیت نهان‌نگاری تقریباً ۱.۵ برابر روش کاراکتر فاصله می‌باشد. لذا روش پیشنهادی در میان روش‌های موجود از بیشترین ظرفیت نهان‌نگاری برخوردار است.

جدول۴. ۱۵ مقایسه ظرفیت نهان‌نگاری با دو روش استفاده از کاراکتر فاصله و روش استفاده هم‌زمان از کاراکتر فاصله و نیم‌فاصله

متن / روش نهان‌نگاری	کاراکتر فاصله	کاراکتر نیم‌فاصله	کاراکتر فاصله و نیم‌فاصله	درصد افزایش ظرفیت نهان‌نگاری درروش ترکیب دو کاراکتر فاصله و نیم‌فاصله به نسبت روش فقط کاراکتر فاصله
متن فناوری	۱۴۱۰	۱۶۸۰	۳۰۹۰	۱۲۰ درصد
متن سیاسی	۶۶۳	۸۷۹	۱۵۴۲	۱۳۲ درصد
متن اقتصادی	۹۲۴	۱۳۲۰	۲۲۲۴	۱۴۲ درصد

۴-۵-۲ امنیت نهان‌نگاری

امنیت نهان‌نگاری را می‌توان در چند سطح تقسیم‌بندی نمود در سطح اول وجود تغییرات ظاهری با تغییرات مشکوک در متن می‌باشد که باعث می‌شود امکان وجود پیام در متن لو برود. در سطح بعدی زمانی لو رفتن پیام یا امکان کشف پیام از طرف دیگران (غیر از گیرنده مشخص) می‌باشد که این مرحله به میزان پیچیدگی الگوریتم‌های رمزنگاری و نهان‌نگاری وابسته است. با توجه به اینکه درروش پیشنهادی تغییری در شکل ظاهری متن ایجاد نمی‌شود لذا امنیت نهان‌نگاری روش پیشنهادی در مقایسه با روش‌هایی که از جابجایی خطوط یا پاراگراف‌ها، تغییر شیب حروف منحنی و تغییر کشیدگی حروف استفاده می‌کنند، بیشتر است. زیرا در این روش‌ها، اولین مرحله امنیت نهان‌نگاری یعنی مشکوک بودن متون به وجود تغییرات در معرض خطر می‌باشد.

البته کشف وجود پیام در این روش نیز امکان‌پذیر است درصورتی‌که متن از یک برنامه کشف تغییرات ویژگی‌های کاراکترها عبور کند آنگاه وجود تغییرات در ویژگی کاراکتر نیم‌فاصله لو می‌رود. اما بازهم امینت این روش نسبت به روش‌های ظاهری بیشتر است چراکه در روش‌های ظاهری از همان ابتدا بارویت متن امکان لو رفتن وجود دارد.

بیشترین امنیت مربوط به روش استفاده از کلمات مترادف می‌باشد در این روش سطح اول امنیت یعنی مشکوک بودن به وجود پیام به‌شدت مطمئن می‌باشد.

برای افزایش امنیت نهان‌نگاری روش پیشنهادی در این پروژه دو کلید تعریف‌شده است. یک کلید مربوط به رمزنگاری می‌باشد که در صورت حتی کشف رشته بیت‌های موجود در متن از طرف شخص ثالث، امکان بازیابی و تبدیل رشته بیت‌ها به کاراکترها حداقل شود. و کلید دیگر برای نهان‌نگاری می‌باشد که درواقع مفاهیم بیت‌های صفر و یک (تغییر یا عدم‌تغییر ویژگی کاراکتر) را در طول متن تغییر می‌دهد. این دو کلید همراه رشته بیت‌های پیام در متن پنهان می‌شوند. روش اجرای آن‌ها در بخش‌های قبلی بیان‌شده است.

در جدول (۴-۱۶) مقایسه امنیت روش پیشنهادی نسبت به سایر روش‌ها آورده شده است. البته این مقایسه به‌صورت کیفی و بر اساس توضیحات بالا می‌باشد.

جدول۴. ۱۶ مقایسه امنیت روش‌های نهان‌نگاری

روش	روش تغییرات ظاهری	روش تغییر ویژگی کاراکتر	روش تغییر ویژگی کاراکتر با تعریف کلید	روش استفاده از کلمات مترادف
میزان امنیت	کم	متوسط	نسبتاً زیاد	زیاد

۴-۵-۳ مقاومت نهان‌نگاری

مقاومت روش نهان‌نگاری بر اساس توانایی حفظ پیام در مقابل تغییرات احتمالی در متن تعریف می‌شود. بدین‌صورت که وجود اشتباه یا تغییر در ویژگی یک یا چند کاراکتر تا چه اندازه بر روی حفظ پیام درست تأثیر می‌گذارد. یکی از روش‌های که برای افزایش مقاومت نهان‌نگاری در کارهای قبلی استفاده‌شده است، تکرار نهان‌نگاری پیام در متن به‌صورت متوالی می‌باشد. که در این حالت با مقایسه پیام‌های متوالی کشف‌شده، احتمال دست‌یابی به پیام درست نسبت به حالت تنها یکبار نهان‌نگاری بسیار بیشتر خواهد بود. با افزایش تعداد دفعات تکرار پیام، احتمال دست‌یابی به پیام درست افزایش‌یافته و درنتیجه مقاومت نهان‌نگاری نیز افزایش می‌یابد. در این پروژه نیز از همین روش برای افزایش مقاومت نهان‌نگاری استفاده‌شده است. البته با توجه به اینکه ظرفیت نهان‌نگاری روش پیشنهادی نسبت به سایر روش‌ها بیشتر است لذا امکان افزایش تعداد دفعات تکرار پیام در یک متن معین برای این روش بسیار بیشتر از سایر روش‌ها می‌باشد لذا می‌توان استدلال نمود که مقاومت

روش پیشنهادی بیشتر از سایر روش‌ها می‌باشد.

می‌توان مقاومت روش نهان‌نگاری به‌صورت رابطه زیر تعریف نمود:

$$\text{مقاومت} = 1 - \frac{k}{n} \quad (۱-۴)$$

که در آن n تعداد کل کاراکترهای متن و k تعداد خطاهای موجود در متن می‌باشد. این خطا می‌تواند تغییر در ویژگی کاراکترها باشد.

درصورتی‌که برای نهان‌نگاری تنها یکی از ویژگی‌های کاراکتر نیم‌فاصله مدنظر باشد لذا رابطه بالا برای تعیین میزان مقاومت روش استفاده می‌شود. اما اگر از چند ویژگی کاراکتر(مثلاً m ویژگی) به‌صورت هم‌زمان استفاده شود، در آن صورت رابطه مقاومت نهان‌نگاری به‌صورت زیر خواهد بود:

$$\text{مقاومت} = 1 - \frac{k}{m * n} \quad (۲-۴)$$

در این حالت کل ویژگی‌های موردبررسی به تعداد $m*n$ خواهد بود لذا درصورتی‌که احتمال وقوع خطا در k ویژگی وجود داشته باشد آنگاه میزان مقاومت به‌صورت رابطه (۴-۲) خواهد شد.

همچنین اگر پیام به تعداد p بار در متن تکرار شود آنگاه احتمال وقوع خطا در پیام برابر با k/p خواهد شد لذا رابطه مقاومت روش نهان‌نگاری باوجود تکرار و استفاده از چند ویژگی به‌صورت زیر خواهد شد:

$$\text{مقاومت} = 1 - \frac{k}{p * m * n} \quad (۳-۴)$$

در اینجا با یک مثال عددی میزان افزایش مقاومت روش نهان‌نگاری برای یکی از متون مورداستفاده در این رساله نشان داده می‌شود. در متن ورزشی که تعداد $n=8929$ کاراکتر وجود دارد اگر احتمال وقوع ۵۰۰ خطا در ویژگی‌ها وجود داشته باشد آنگاه مقاومت در سه حالت ذکرشده به‌صورت زیر خواهد بود:

حالت اول- نهان‌نگاری به کمک یک ویژگی و بدون تکرار

$$\text{مقاومت} = \left(1 - \frac{500}{8929}\right) * 100 = 94.4\%$$

حالت دوم- نهان‌نگاری به کمک ۵ ویژگی و بدون تکرار

$$\text{مقاومت} = \left(1 - \frac{500}{5 * 8929}\right) * 100 = 98.88\%$$

حالت سوم- نهان‌نگاری به کمک ۵ ویژگی و ۵ تکرار

$$\text{مقاومت} = \left(1 - \frac{500}{5 * 5 * 8929}\right) * 100 = 99.77\%$$

همان‌طور که مشاهده می‌گردد با افزایش تعداد ویژگی‌های مورداستفاده و تعداد تکرار پیام در متن، مقاومت روش نهان‌نگاری به مقدار قابل‌توجهی افزایش می‌یابد به‌طوری‌که مقاومت به عدد ۱۰۰ درصد نزدیک می‌شود یعنی روش نهان‌نگاری در مقابل تغییرات بسیار مقاوم خواهد بود.

ازآنجایی‌که تعداد ویژگی‌های کاراکترها محدود می‌باشد لذا برای دست یافتن به مقاومت بسیار بالا می‌توان تعداد تکرار پیام را در متن افزایش داد. ازآنجایی‌که تعداد تکرار پیام در یک متن خاص وابسته به میزان ظرفیت نهان‌نگاری متن می‌باشد لذا هرچقدر ظرفیت نهان‌نگاری بالاتر باشد به‌صورت مستقیم تعداد تکرار پیام در متن و درنتیجه مقاومت روش نهان‌نگاری افزایش خواهد یافت.

فصل ۵ : نتیجه‌گیری و پیشنهادات

۵-۱ نتیجه‌گیری

در این پروژه روشی جدید برای نهان‌نگاری متون فارسی مبتنی بر استفاده از ویژگی‌های کاراکترهای فاصله و نیم‌فاصله ارائه شده است. با توجه به اینکه در حالت معمول تعدد کاراکتر نیم‌فاصله در متون فارسی کم می‌باشد لذا با ایجاد کاراکتر نیم‌فاصله مابین حروف غیرچسبان، فرکانس تکرار این کاراکتر در متون فارسی در این پروژه به‌شدت افزایش‌یافته است طوری که در این حالت تعدد کاراکتر نیم‌فاصله از کاراکتر فاصله نیز بیشتر شده است. و درنهایت درروش پیشنهادی با ترکیب دو کاراکتر فاصله و نیم‌فاصله، میزان ظرفیت روش پیشنهادی به بیش از ۱.۳۱ برابر روش‌های معمول افزایش‌یافته است. علاوه بر شاخص ظرفیت نهان‌نگاری، شاخص امنیت نیز بهبود داده‌شده است. در این پروژه با تعریف کلیدهای رمزنگاری و نهان‌نگاری احتمال کشف پیام از سوی شخص ثالث نسبت به سایر روش‌ها کاهش‌یافته است. کلیدهای رمزنگاری نحوه تبدیل کاراکترها به کدهای باینری را تعیین می‌کند که حالت‌های مختلفی دارند و کلیدهای الگوریتم مفاهیم بیت‌های صفر و یک را در طول متن تغییر می‌دهند لذا با این اقدام امنیت نهان‌نگاری به روش پیشنهادی نسبت به روش‌های معمول افزایش‌یافته است. و درنهایت شاخص مقاومت نهان‌نگاری به‌صورت غیرمستقیم بهبودیافته است که درواقع این بهبود تحت تأثیر افزایش ظرفیت نهان‌نگاری می‌باشد با توجه به اینکه ظرفیت نهان‌نگاری درروش پیشنهادی بیشتر از سایر روش‌ها می‌باشد لذا امکان تکرار بیشتر پیام در متن معین نسبت به سایر روش‌ها وجود دارد و ازآنجایی‌که تعداد دفعات تکرار پیام در متن باعث افزایش مقاومت نهان‌نگاری می‌شود لذا مقاومت نهان‌نگاری به روش پیشنهادی بیشتر از سایر روش‌ها می‌باشد.

۵-۲ پیشنهادات

در این بخش پیشنهاداتی برای اجرای نهان‌نگاری به روش‌های دیگر ارائه شده است

۱- استفاده ترکیبی از روش‌های نهان‌نگاری، استفاده از کلمات مترادف، تغییر شکل ظاهری حروف و استفاده از تغییر ویژگی‌های کاراکترهای فاصله و نیم‌فاصله به‌منظور افزایش ظرفیت و امنیت نهان‌نگاری

۲- استفاده از ویژگی‌های هجاها و حروف صامت و مصوت در زبان فارسی و نحوه قرارگیری آن‌ها در هجاها به‌منظور کاهش رشته بیت موردنیاز در مرحله رمزنگاری

۳- بررسی نهان‌نگاری به کمک تعداد کاراکترهای نیم‌فاصله بین حروف غیرچسبان، ازآنجایی‌که تعداد کاراکترهای نیم‌فاصله تغییری در متن ایجاد نمی‌کند لذا می‌توان به هر تعداد کاراکتر نیم‌فاصله بین حروف غیرچسبان اضافه نمود.

منابع :

[1] S.Katzenbeisser and F.A.P.Petitcolas, (2000), "**Information hiding techniques for steganography and digital watermarking**", Artech House, London.

[2] C.E.Shannon, (Oct 1949), "**Communication Theory of Secrecy Systems**", Bell system technical journal vd . 28-4 .

[3] E.Cole, (2003), "**Hiding in plain sight: steganography and the art of covert communication**", Wiley Publishing Inc.

[4] G.Kipper, (2004), "**Investigator's guide to steganography**", Auerbach Publications.

[5] D. Holmes, (February 2002), "**Introduction to digital image steganography**", Available in: http://www.csie.mcu.edu.tw/~s9170464/steganography.doc.

[6] S.Katzenbeisser and F.A.P.Petitcolas, (2002), "**Defining security in steganographic systems**" Proceedings of the SPIE vol. 4675, Security and Watermarking of Multimedia Contents IV, pp. 50-56.

[7] N. Johnson, Z. Duric and S. Jajodia,(May 2001), "**Information hiding: steganography and watermarking-attacks and countermeasures**", Kluwer Academic Publishers, Boston, pp. 85-86.

[8] A.Sarkar, K.Solanki and B.S.Manjunath, (2008),"**Further study on YASS : steganography based on randomized embedding to resist blind steganalysis**" Proc. of SPIE Security, Steganography, and Watermarking of Multimedia Contents.

[9]Birgit Pfitzman, (1996), "**Information Hiding Terminology**", Lecture Notes in computer science 1174, Springer, pp. 358-347

[10]Shawn D.Dickman,(2007),**"An Overview of Steganography"**, James Madison University infosec Techreport.

[11] I_ J. Cox, G. Doerr, and T. Furon, (2006), "**Watermarking is not cryptography**" IWDW'06, Springer- Verlag, pp.

[12] Ammar Odeh and Khaled Elleithy, (June 2012), "**Steganography in Arabic Text Using Zero Width and Kashidha Letters**", International Journal of Computer Science & Information Technology (IJCSIT), Vol 4, No 3, pp. 1 – 11.

[13] Adnan Gutub and Manal Fattani, (May 2007), "**A Novel Arabic Text Steganography Method Using Letter Points and Extensions**", WASET International

Conference on Computer, Information and Systems Science and Engineering (ICCISSE), Vienna, Austria.

[14] Adnan M. Alattar and Osama M. Alattar,(2004), "**Watermarking Electronic Text Documents Containing Justified Paragraphs and Irregular Line Spacing**", Digimarc Corporation, Tualatin, OR 97062

[15]J. T. Brassil, S. Low, N. F. Maxemchuk, and L. O'Gorman, (October 1995), "**Electronic Marking and Identification Techniques to Discourage Document Copying**", IEEE Journal on Selected Areas in Communications, vol. 13, no. 8, pp. 1495-1504.

[16] J. T. Brassil, S. lnw, and N. F Maxemchuk, (July 1999) , "**Copyright Protection for the Elechonic Distribution of Text Documents**" , Proceedings of the IEEE, vol. 87, no.7, pp.1181 – 1196 .

[17] S. H. Low, N.F. Maxemchuk, J.T. Brassil, and L.O'Gorman, (April 1995) , "**Document Marking and Identification Using Both Line and Word Shifting**", Proc. Infoncom'95, Boston, MA, pp. 853-860.

[18] M.H. Shirali-Shahreza and M. Shirali-Shahreza,(2006), "**A New Approach to Persian/Arabic Text Steganography**", Proceedings of the 5th 6 IEEE/ACIS International Conference on Computer and Information Science (ICIS 2006), Honolulu, HI, USA, pp. 310- 315, July 10-12.

[19] Adnan Gutub, Lahouari Ghouti, Alaaeldin Amin, Talal Alkharobi and Mohammad K.Ibrahim , (July 2007) , "**Utilizing Extension Character 'Kashida' With Pointed Letters For Arabic Text Digital Watermarking**" International Conference on Security and Cryptography–(SECRYPT),Barcelona, Spain.

[20] MH Shirali-Shahreza and S Shirali-Shahreza, (2005) , "**A Robust Page Segmentation Method for Persian/Arabic Document**" , WSEAS Transactions on Computers .

[21] Mohit Garg, (October 2011) , " **A Novel Text Steganography Technique Based on Html Documents**" ,International Journal of Advanced Science and Technology,pp.132-138,Vol. 35.

[22] MH Shirali-Shahreza and S Shirali-Shahreza , (2008) , "**Steganography in TeX documents**"

Intelligent System and Knowledge Engineering .ISKE 2008. 3rd International Conference .

[23] M Shirali-Shahreza, (2007) , "**A New Persian/Arabic Text Steganography Using "La" Word**" , Proceedings of the International Joint Conference on Computer, Information, and Systems Sciences, and Engineering (CISSE 2007), Bridgeport, CT, USA,Vol. 2, pp. 339-342.

[24] Vahid Yazdani , Mohammad Ali Doostari and Hamid Yazdani , (2013) ," **A New Method to Persian Text Watermarking Using Curvaceous Letters** " **,** J. Basic Appl. Sci.Res , pp.125-131.

[25] patel and monika , (2012) , " **Analytical Study of Line-Shift Text watermarking Technique**" **,** International Journal of Computer Applications and Information Technology , pp.84-87 .

[۲۶]ترشیزی ع ، (۱۳۹۳) ، پایان‌نامه کارشناسی ارشد: **" نهان‌نگاری اطلاعات در متون فارسی** " ، دانشکده فنی و مهندسی گروه کامپیوتر ، دانشگاه آزاد اسلامی شاهرود .

Abstract

Creation and Development of a secure platform for digital communication is one of the initial and basic requirements of digital world. With expansion of the Internet functionality along with rapid growth on digital data technology, the need is increasing for a secure communication platform, where data can be exchanged securely. Providing a secure communication between transmitter and receiver is objective.

There are several means to achieve this goal. Encryption is one of the methods which can be utilized. In encryption, process of encoding messages or information can be carried out in such a way that only authorized parties can read it. Steganography is the other method in providing secure communication. The practice of concealing messages or information within other non-secret text or data is called Steganography.

In this paper we use pseudo space character and normal space character for hiding bits.in Arabic alphabet there are both non joiner letters and joiner letters, pseudo characters are used in non-joiner letters to increase capacity of Steganography up to 1.31 times of conventional method. Security of steganography can be boosted by variation in used algorithm with changing of interpretation of bits either one or zero. To increase the resistance of Steganography, message should be repeated in text several times, just one after the other. This will ensure that in extraction sequence all errors can be detected with comparing of extracts and correct message can be obtained. In addition, resistance of steganography can be enhanced by combining of different properties of pseudo character e.g. (type, size, color, boldness and italic shape). In this mode all properties will be compared to each other to be able to extract correct message without any error. Considering all above, these two methods will ensure up to 99.77% increase in resistance of steganography which is depicting high resistance of this method.

Keywords: Steganography, Pseudo space Character, Steganography resistance, Steganography capacity, Steganography security.

Shahrood University of Technology

Faculty E-learning center

Steganography in Persian text using statistical methods

Hajar Nadimi

Supervisor:

Morteza Zahedi

Advisor:

Aliakbar Poyan

Date: **February 2016**