نشریه علمی «علوم و فناوری‌های پدافند نوین»
سال دوازدهم، شماره ۲، تابستان ۱۴۰۰؛ ص ۱۶۷- ۱۵۹

علمی- پژوهشی

پنهان‌نگاری در متن بر اساس روش‌های ترازبندی متن

بهمن خسروی[۱*]، بهنام خسروی[۲]
۱- استادیار دانشگاه صنعتی قم، ۲- استادیار دانشگاه تحصیلات تکمیلی زنجان
(دریافت: ۱۳۹۸/۱۰/۱۴، پذیرش: ۱۴۰۰/۰۲/۰۴)

چکیده

یکی از مهم‌ترین فن‌ها در حفظ اطلاعات، پنهان کردن اطلاعات است. پنهان‌نگاری اطلاعات علم و هنر پنهان کردن اطلاعات در لابه‌لای سایر داده‌هاست (به‌عنوان‌مثال متن، تصویر، فیلم یا صدا)، به‌گونه‌ای که یافتن اطلاعات یا پی بردن به‌وجود آن سخت یا حتی غیرممکن باشد. در این مقاله با استفاده از روش‌هایی که برای ترازبندی (منظم‌سازی) متن در نرم‌افزارهای تایپ استفاده می‌شوند، روشی برای پنهان‌نگاری اطلاعات در متن ارائه شده است. روش ارائه شده در این مقاله توانایی درج اطلاعات بیشتری نسبت به برخی الگوریتم‌های پیشین در این زمینه دارد. این الگوریتم در برابر چند نوع از حملات ازجمله حملات بصری، ساختاری، آماری و ... مقاوم است. قابلیت حائز اهمیت دیگر این روش قابل چاپ بودن آن است که می‌توان برای ارسال اطلاعات از نسخه چاپ شده آن روی کاغذ هم استفاده کرد.

کلیدواژه‌ها: پنهان‌نگاری، متن‌های ترازبندی شده، پنهان‌نگاری در PDF، رمزنگاری

Text Steganography Based on Text Justifying Methods

B. Khosravi*, B. Khosravi
Qom University of Technology, Qom
(Received: 04/01/2020; Accepted: 24/04/2021)

Abstract

One of the most important information security techniques is the hiding of information. Steganography is the art and science of hiding information in the cover of data (in the form of text, image, video, or audio) such that it does not arise any suspicions, and is difficult or even impossible to discover. This paper presents a method for steganography in the form of text which uses the methods of text justifying in typing editors. The method presented in this paper is able to hide information better than some of previous algorithms in this field. This algorithm is resistant to various forms of attack such as visual, structural and statistical attacks. Another important capability of this method is that it can be used to send printed information.

Keywords: Steganography, Hiding Information, Justified Text, PDF Steganography, Cryptography.

*Corresponding Author E-mail:khosravi@qut.ac.ir

۱- مقدمه

امنیت ارتباطات بین دو نفر همیشه یکی از مهم‌ترین مسائل و دغدغه‌ها بوده است. با ظهور اینترنت، نگرانی از ارتباطات امن افزایش یافته است، به‌گونه‌ای که امروزه امنیت یکی از مهم‌ترین بخش‌ها در فناوری اطلاعات و ارتباطات است. بدین منظور رمزنگاری به‌عنوان وسیله‌ای برای ایمن نگه‌داشتن ارتباطات ایجاد شده است، اما رمزنگاری همیشه برای حفظ امنیت ارتباطات کافی نیست زیرا متن رمزگذاری شده توسط شخص ثالث دیده می‌شود و ممکن است رمز شکسته شود یا از انتقال داده‌های رمزگذاری شده جلوگیری شود؛ بنابراین علم پنهان‌نگاری به‌وجود آمد.

پنهان‌نگاری (استگانوگرافی) هنر و علم برقراری ارتباط به‌گونه‌ای است که نتوان وجود یک پیام را تشخیص داد [۱]. کلمه استگانوگرافی از کلمات یونانی"stegos" به معنای «پوشیده» و "grafia" به معنای «نوشتن» اخذ شده است که آن را «پوشیده نوشتن» می‌توان معنا کرد [۱]. روش‌های متعددی در طول تاریخ برای پنهان‌نگاری استفاده شده‌اند، اما با توسعه پَروَنجاهای[1] الکترونیکی، روش‌ها و فنون متعدد جدیدی ارائه شده‌اند.

چهار داده‌ای که معمولاً برای مخفی کردن اطلاعات از آن‌ها استفاده می‌شوند عبارت‌اند از متن، تصویر، صدا و فیلم. ازاین‌رو پنهان‌نگاری را عموماً به چهار دسته طبقه‌بندی می‌کنند، یعنی پنهان‌نگاری متن، پنهان‌نگاری تصویر، پنهان‌نگاری صوتی و پنهان‌نگاری ویدیویی. با توجه به استفاده گسترده متون در زندگی روزمره، پنهان‌نگاری در متن از اهمیت به سزایی برخوردار است ولی با توجه به این‌که تغییرات کمی می‌توان در متن داد که باعث ایجاد شک و ظن نشوند، پنهان‌نگاری در متن به‌نظر می‌رسد ازجمله سخت‌ترین انواع پنهان‌نگاری‌ها باشد.

روش‌های متعددی برای انواع مختلف پنهان‌نگاری از جمله پنهان‌نگاری در متن ارائه شده‌اند. از این رو می‌توان به روش فشرده‌سازی نمایه‌سازی [۲]، روش ردیابی تغییر [۳]، قرار دادن اطلاعات در پویانمایی یک پَروَنجا پاورپوینت [۴]، درج پیام با تغییر رنگِ کاراکترهای نامرئی [۵]، ویژگی‌های زبان [۸- ۶] و ... اشاره نمود. یکی از روش‌های رایج در روش‌های پنهان‌نگاری در متن استفاده از فاصله‌هایی با طول صفر است که در لابه‌لای کلمات و کاراکترها اضافه می‌شوند. می‌دانیم برخی از نرم‌افزارهای واژه‌پرداز مثل ورد به‌راحتی این کاراکترهای نامرئی را نشان می‌دهند، همچنین با اضافه کردن این کاراکترها حجم پَروَنجا افزایش می‌یابد و لذا به‌راحتی مشخص می‌شود که متن حاوی پیام پنهانی است. از طرف دیگر این روش‌ها قابل‌استفاده در متن‌های چاپ شده نیستند [۹].

نویسه‌خوانی نوری را به‌اختصار با OCR نمایش می‌دهند. واژه OCR مخفف Optical Character Recognition و یا Optical Character Reader است. نویسه‌خوانی نوری فنّاوری است که به کمک آن می‌توان صفحات حاوی متون اسکن شده را از فرمت عکس به متن تبدیل کرد.

تراز کردن متون (مرتب‌سازی متون) یکی از توانمندی‌های اساسی در نرم‌افزارهای واژه‌پرداز است که باعث آراستگی متون می‌شود. روش ترازبندی متون بدین‌صورت است که با استفاده از اضافه کردن فاصله اضافی بین برخی از کاراکترها و کشیدن یا فشرده‌سازی برخی کاراکترها، لبه‌های ناقص را از چپ و راست متن حذف می‌کند. درصورتی‌که فاصله بین دو کاراکتر از حدی بیشتر کشیده شود، زمانی که با برخی از نرم‌افزارهای نویسه‌خوانی نوری متن خوانده شود به‌جای نمایش یک ۳۲ که، کد اسکی فاصله است، دو تا ۳۲ نمایش داده می‌شود. دراینجا این فاصله‌ها، فاصله اضافه شده نامیده می‌شود. در ادامه فاصله‌های اضافه شده را با AS و فاصله‌های معمولی را با S نمایش می‌دهیم.

اگر خطی در متن با حداقل ۱۳ فاصله وجود داشت به‌طوری‌که حداقل دارای ۹ فاصله معمولی (S) و حداقل دارای ۳ فاصله اضافه‌شده (AS) باشد، به آن خط میزبان می‌گویند و آن با استفاده از HL، نمایش داده می‌شود. چنانچه خطی دارای ۱۴ فاصله باشد که حداقل دارای ۹ فاصله معمولی (S) و حداقل دارای ۴ فاصله اضافه شده (AS) باشد آن خط میزبان فراوانی‌ها، نامیده می‌شود و با استفاده از FHL، نمایش داده می‌شود.

الگوریتم ارائه شده در این مقاله با تغییر مکان AS در خط، اطلاعات را پنهان می‌کند. این الگوریتم اولین بار توسط بهروز خسروی و دیگران ارائه شد [۱۰] و نسبت ظرفیت ارائه شده در آن بسیار بهتر از برخی از الگوریتم‌های موجود بود. این در حالی است که از امنیت بسیار مناسبی نسبت به برخی از الگوریتم‌ها برخوردار است. با روش ارائه شده در آن مقاله در هر خط میزبان می‌توان ۴ بیت درج کرد [۱۰]. در این مقاله روش جدیدی ارائه شده است که به کمک آن می‌توان در هر خط میزبان ۶ بیت درج کرد و لذا می‌توان یک و نیم برابر اطلاعات در متن پنهان نمود.

مقاله به‌صورت زیر سازمان‌دهی شده است: در بخش دوم، روش جاسازی اطلاعات آورده شده است که در آن الگوریتم درج و الگوریتم پنهان‌سازی ارائه شده‌اند. روش بازیابی اطلاعات در بخش سوم ارائه شده است. در بخش چهارم نتایج تجربی و تجزیه‌وتحلیل الگوریتم آورده شده‌اند. در نهایت نتیجه‌گیری در بخش پنجم ارائه شده است.

[1] File

۲- یک روش جدید برای جاسازی اطلاعات

تعداد تمام شش بیتی‌ها برابر با ۶۴ = $۲^۶$ است و از طرفی تعداد تمام درایه‌های یک ماتریس سه‌بعدی A (ماتریس مکعبی) نیز ۶۴ = ۴ × ۴ × ۴ است؛ بنابراین می‌توان هر ۶ بیتی را به یک درایه ماتریس ۴ × ۴ × ۴ نظیر کرد، به‌عبارت‌دیگر می‌توان یک ماتریس ۴ × ۴ × ۴ را در نظر گرفت و هر ۶ بیتی را در یکی از درایه‌های آن قرار داد. حال این ماتریس حاصل‌شده از این به بعد ثابت نگه داشته و سپس بین فرستنده و گیرنده به اشتراک گذاشته می‌شود. در ادامه روش جدیدی برای درج اطلاعات و پنهان‌نگاری ارائه می‌شود.

۲-۱- الگوریتم درج

همان‌طور که در بالا اشاره شد می‌توان هر ۶ بیتی را با یک درایه از ماتریس ۴ × ۴ × ۴ متناظر کرد. درنتیجه هر ۶ بیتی توسط یک سه‌تایی مرتب (i , j , k) که $۰ \le i , j , k \le ۳$ نشان داده می‌شود. درنتیجه به‌جای جاسازی کردن یک ۶ بیتی می‌توان سه‌تایی مرتب نظیر آن جاسازی شود. به بخشی از یک خط که شامل ۴ کلمه متوالی باشد و لذا شامل ۳ فاصله است، از این پس یک بلوک نامیده می‌شود. در ادامه روشی ارائه می‌شود، تا بتوان اعداد ۰، ۱، ۲ و ۳ را در یک بلوک درج کرد.

حال به‌منظور جاسازی کردن سه‌تایی مرتب (i , j , k) که $۰ \le i , j , k \le ۳$ در یک HL از روش زیر استفاده می‌شود.

- اگر قصد داشته 0 در یک بلوک جاسازی گردد، با جابه‌جا کردن محل S و ASها در خط، تمام سه فاصله موجود در آن بلوک به فاصله معمولی (S) تبدیل می‌شود.
- برای درج ۱، ۲ یا ۳ در یک بلوک به‌صورت زیر عمل می‌شود، برای جاسازی ۱ اولین فاصله در صورت لزوم به فاصله اضافه شده (AS) تغییر می‌یابد و دو فاصله دیگر اگر فاصله معمولی نباشند به فاصله معمولی (S) تغییر داده می‌شود. به‌طور مشابه برای درج ۲ و ۳ به‌ترتیب دومین و سومین فاصله در بلوک به یک‌فاصله اضافه‌شده (AS) تغییر داده می‌شود و دو فاصله دیگر می‌بایست فاصله معمولی (S) باشند.

فرض شود که تعداد فاصله‌های موجود در یک HL (شامل فاصله‌های (S) و فاصله‌های (AS)) برابر l باشد؛ حال در این HL سه بلوک، بدین شکل در نظر گرفته می‌شود.

- چهار کلمه اول خط، به‌عنوان اولین بلوک در نظر گرفته می‌شود. لذا اولین بلوک دارای فاصله‌های اول، دوم و سوم در خط است.
- دومین بلوک، طوری انتخاب می‌شود که، دارای سه فاصله میانی در خط باشد و دو حالت رخ می‌دهد. اگر l عددی فرد باشد فاصله‌های موجود در مکان‌های $\frac{l-1}{2}$، $\frac{l+1}{2}$ و $1+\frac{l+1}{2}$ در نظر گرفته می‌شود و درصورتی‌که l عددی زوج باشد فاصله‌های موجود در مکان‌های $\frac{l}{2}-1$، $\frac{l}{2}$ و $1+\frac{l}{2}$ در نظر گرفته می‌شود.
- سومین بلوک شامل فاصله‌های موجود در مکان‌های $l-۳$ ، $l-۲$ و $l-۱$ است (این بلوک آخرین فاصله را شامل نمی‌شود).

توجه شود که هر دو بیتی به‌عنوان عددی در مبنای ۲، عددی بین ۰ تا ۳ در مبنای ۱۰ است. برای درج یک ۶ بیتی آن را به سه، ۲ بیتی می‌توان تقسیم کرد و هر دو بیت را به‌عنوان عددی در مبنای ۲ در نظر گرفت و آن را به مبنای ۱۰ برد و سپس عدد نظیر اولین دو بیتی، در اولین بلوک خط میزبان درج می‌شود و به همین ترتیب دومین و سومین اعداد در بلوک‌های دوم و سوم خط میزبان درج می‌شوند.

بنابر تعریف خط میزبان، در یک HL حداقل چهار فاصله دیگر باقی‌مانده است. همچنین بنا بر تعریف HL حداقل سه تا از فاصله‌های آن از نوع AS است. در بدترین حالت فرض می‌شود که هیچ‌کدام از ASها درج نشده باشند (یعنی سه‌تایی مرتب $(0 , 0 , 0)$ درج‌شده باشد و لذا در هر ۳ بلوک فاصله‌ها از نوع معمولی‌اند)؛ بنابراین سه AS در ۳ فاصله باقی‌مانده (به‌جز آخرین فاصله در خط) درج می‌شود؛ یعنی در سه تا از فاصله‌های باقی‌مانده در HL (که به بلوکی تعلق ندارند و آخرین فاصله نیز نیستند) AS های باقی‌مانده در صورت وجود درج می‌شود.

۲-۲- الگوریتم پنهان‌سازی

حال در این بخش با استفاده از روش جاسازی فوق الگوریتمی برای پنهان‌سازی، ارائه می‌گردد.

گام۱: متن پیام سری، با استفاده از کدگذاری هافمن، فشرده‌سازی می‌گردد. رشته (دودویی) حاصل‌شده را T می‌نامند.

گام۲: به‌منظور افزایش امنیت رشته حاصل‌شده از گام قبل (یعنی T)، به بلوک‌های هفت‌تایی افراز می‌گردد. سپس یک بیت به انتهای هر بلوک اضافه می‌شود و جمع XOR هر بلوک هفت‌تایی درون این بیت اضافه شده قرار داده می‌شود. بنابراین طول هر بلوک هشت می‌شود و درمجموع رشته دودویی جدیدی حاصل می‌شود که به آن B گویند.

گام۳: برای افزایش امنیت، الگوریتم رمزنگاری زیر نیز استفاده می‌شود. بدین منظور یک ماتریس ۶ × ۶ با درایه‌های ۰ یا ۱ در نظر گرفته می‌شود که دترمینان آن عددی فرد است (اگر

دترمینان ماتریسی فرد باشد در میدان گالوای از مرتبه ۲ یعنی (۲) GF دارای وارون است). حال به‌ترتیب هر خط ماتریس که ۶ بیت است، در یک خط HL نشانده می‌شود. بنابراین شش سطر ماتریس در اولین شش خط میزبان درج می‌گردند. حال رشته دودویی B به بلوک‌های شش‌تایی افراز می‌شود. اگر طول رشته مضربی از شش نباشد، با اضافه کردن صفر به ابتدای رشته (از سمت چپ) طول رشته به مضربی از شش رسانده می‌شود. سپس هر بلوک در ماتریس مزبور ضرب می‌شود (عمل ضرب در میدان GF(۲) است و لذا ماتریس حاصل نیز درایه‌هایش ۰ و ۱ هستند). بنابراین بلوک‌های شش‌تایی جدیدی حاصل می‌شود که، رمزگذاری شده‌اند. رشته حاصل‌شده جدید، Z درنظر گرفته می‌شود.

گام۴: **جاسازی فراوانی کاراکترها:** فرستنده و گیرنده، لیست و ترتیب کاراکترها را از قبل باهم توافق کرده‌اند. برای بازیابی اطلاعاتی که با استفاده از کد هافمن کدگذاری شده‌اند، گیرنده نیازمند دانستن فراوانی کاراکترهاست. بدین منظور ابتدا فراوانی تمام کاراکترها را در مبنای دو می‌نویسند. اگر طول رشته‌ی مربوط به فراوانی کاراکتری مضربی از شش نباشد با افزودن تعدادی صفر به سمت چپ آن، طول رشته‌ی مورد نظر به مضربی از شش رسانده می‌شود.فراوانی اولین کاراکتر در FHL بعدی (که هفتمین خط میزبان از نوع خط میزبان فراوانی است) درج می‌شود. برای این منظور اگر طول رشته مربوط به فراوانی اولین کاراکتر در مبنای دو مساوی شش باشد، با استفاده از الگوریتم فوق این رشته در خط میزبان درج می‌شود و آخرین فاصله در آن خط میزبان در صورت لزوم به S تبدیل می‌شود (اگر AS بود به S تبدیل می‌شود). درصورتی‌که طول فراوانی کاراکتر موردنظر از شش بیشتر بود، به بلوک‌هایی با طول شش افراز می‌گردد. اولین بلوک این رشته در اولین FHL نشانده می‌شود و آخرین فاصله خط در صورت لزوم به AS تبدیل می‌گردد تا نشان دهد که خط بعدی به این خط وابسته است و این فرآیند آن‌قدر ادامه می‌یابد تا کل آن نشانده شود. سپس در آخرین FHL استفاده‌شده، آخرین فاصله‌اش در صورت لزوم به S تبدیل می‌شود تا نشان دهد که درج فراوانی این کاراکتر تمام شده است. این فرآیند تا جایی ادامه می‌یابد که فراوانی کل کاراکترها درج شود.

گام۵: **نشاندن متن پیام:** رشته Z حاصل‌شده در گام ۳ به بلوک‌های به طول شش افراز می‌گردد و هر بلوک، با استفاده از الگوریتم فوق در یک خط میزبان (HL) درج می‌گردد و این فرآیند آن‌قدر ادامه می‌یابد تا کل رشته Z درج شود. درنهایت در خط میزبان بعدی، اولین و سومین فاصله به AS تبدیل می‌شود تا با خطوط دیگر متفاوت باشد و نشان دهد که فرآیند نشاندن به اتمام رسیده است. این خط را خط پایانی می‌نامند.

۳- روش بازیابی اطلاعات

به‌منظور بازیابی اطلاعات از فرآیند زیر استفاده می‌شود. پَروَنجا PDF یا متنی که اطلاعات در آن ذخیره شده‌اند، با نرم‌افزار نویسه‌خوانی نوری خوانده می‌شود و خطوط میزبان (HLها)، در آن مشخص می‌گردد. از شش خط میزبان اول، M کلید رمز، استخراج می‌گردد و بعد از مشخص شدن کلید M وارون آن، M^{-1} ، مشخص می‌شود. پس‌از آن به آخرین فاصله در هفتمین خط میزبان، توجه می‌شود. اگر S بود فراوانی اولین کاراکتر از آن استخراج می‌شود ولی اگر AS بود رشته استخراج‌شده از خط میزبان بعدی به رشته مستخرج از هفتمین خط میزبان الحاق می‌شود و این فرآیند ادامه پیدا می‌کند تا به خط میزبانی برسد که آخرین فاصله‌اش S باشد. حال فراوانی اولین کاراکتر مشخص می‌شود. همین فرآیند تا جایی ادامه می‌یابد که فراوانی کل کاراکترها استخراج شود. ازآنجایی‌که تعداد کاراکترها مشخص است، بنابراین پس از به‌دست آوردن فراوانی آخرین کاراکتر معلوم است که خط میزبان بعدی حاوی متن اصلی است. فرآیند استخراج آن‌قدر ادامه می‌یابد تا به خط پایانی برسد.

۳-۱- الگوریتم بازیابی

برای بازیابی اطلاعات از الگوریتم زیر استفاده می‌شود. ابتدا متن با نرم‌افزار نویسه‌خوانی نوری خوانده می‌شود و خطوط میزبان (HL) در آن مشخص می‌گردد.

در یک خط میزبان (HL) سه بلوک مشخص می‌شود. اولین بلوک شامل سه فاصله اول است، دومین بلوک شامل سه فاصله میانی و بلوک سوم شامل سه فاصله انتهایی خط به‌جز آخرین فاصله است. در خطوط میزبانی که خط پایانی نیستند، سه‌تایی مرتب (i , j , k) نظیر این سه بلوک، به‌صورت زیر به‌دست می‌آید.

- اگر در بلوک هیچ AS ای وجود نداشت، t = ۰ ؛
- اگر اولین فاصله در بلوک، AS بود، t = ۱ ؛
- اگر دومین فاصله در بلوک، AS بود، t = ۲ ؛
- اگر سومین فاصله در بلوک، AS بود، t = ۳ ؛

اگر اولین و سومین بلوک AS بود، خط پایانی است و فرآیند استخراج به اتمام رسیده است.

از شش خط اول کلید M مشخص می‌شود و وارون آن به‌دست می‌آید. همان‌طور که در بالا گفته شد فراوانی کاراکترها و متن پیام ذخیره شده Z استخراج می‌گردد. حال رشته استخراج‌شده به

بلوک‌های به طول شش افراز شده و هرکدام در M^{-1} ضرب می‌شود. درنهایت این شش‌تایی‌های جدید به‌ترتیب به یکدیگر چسبانده می‌شود تا رشته B به‌دست آید. سپس این رشته به بلوک‌های به طول هشت افراز می‌گردد. اگر جمع XOR هفت بیت اول برابر هشتمین بیت شد، نشان می‌دهد که هفت بیت اول دچار اختلال نشده‌اند. این فرآیند تا بررسی آخرین بلوک هشت‌تایی ادامه می‌یابد. سپس هفت بیت اول هرکدام از بلوک‌ها به یکدیگر چسبانده می‌شود، تا رشته اصلی T حاصل شود. درنهایت به کمک فراوانی کاراکترها و الگوریتم کدگشایی کد هافمن، متن پیام به‌دست می‌آید.

۴- نتایج تجربی و تجزیه و تحلیل الگوریتم

در این بخش ابتدا با ارائه یک مثال، نحوه پیاده‌سازی الگوریتم مورد بررسی قرار می‌گیرد و سپس به تجزیه و تحلیل آن پرداخته می‌شود.

۴-۱- پیاده‌سازی الگوریتم و ارائه یک مثال

برای پنهان کردن کلمه steganography در متن پوششی به صورت زیر عمل می‌شود.

گام ۱: کد هافمن متناظر با کلمه steganography برابر است با:
T=11101010011010011001110101001011110000001111 1

طول این رشته برابر ۴۵ است. همان‌طور که پیش‌ازاین نیز اشاره شد، برای بازیابی کد هافمن باید جدول فراوانی کاراکترها موجود باشد. فروانی کاراکترهای کلمه steganography در جدول (۱) ارائه شده‌اند. توجه شود که فراوانی کاراکترهای دیگری که در این کلمه ظاهرنشده‌اند برابر صفر است.

جدول۱. جدول فراوانی کاراکترهای آن عبارت است از:

کد تخصیص داده شده	فراوانی	کاراکتر
۱۰۰	۲	g
۱۱۰	۲	a
۰۱۰	۱	o
۰۰۰	۱	p
۰۰۱	۱	h
۰۱۱۱	۱	n
۱۱۱۰	۱	s
۱۰۱۱	۱	r
۰۱۱۰	۱	e
۱۰۱۰	۱	t
۱۱۱۱	۱	y

گام ۲: در گام دوم رشته T به بلوک‌های به طول هفت افراز می‌شود. توجه شود که طول رشته T برابر با ۴۵ است، پس شش بلوک هفت‌تایی از آن جدا می‌شود و بیت هشتم به هر یک از آنان مطابق الگوریتم اضافه می‌گردد (توجه شود که سه کاراکتر آخر در این مرحله تأثیری ندارند). بنابراین طول رشته حاصل‌شده در این مرحله ۵۱ خواهد بود.

B =
111010100110100110011101010010111100000011111

گام ۳ (رمزگذاری): به عنوان مثال فرض شود که ماتریس کلید به‌صورت زیر باشد. طول رشته B برابر ۵۱ است. برای آن‌که طول رشته مضربی از شش شود، به تعداد سه بیت صفر به سمت چپ آن اضافه می‌شود تا طول آن به ۵۴ برسد. حال به بلوک‌های به طول شش افراز شده و هر بلوک در ماتریس کلید ضرب می‌شود.

$$M=\begin{bmatrix}1&0&0&1&0&0\\0&1&0&1&1&0\\1&1&0&1&1&0\\0&0&0&0&1&0\\1&1&1&1&1&1\\0&1&1&1&1&0\end{bmatrix}$$

بنابراین رشته زیر به دست می‌آید که طولش برابر ۵۴ است.

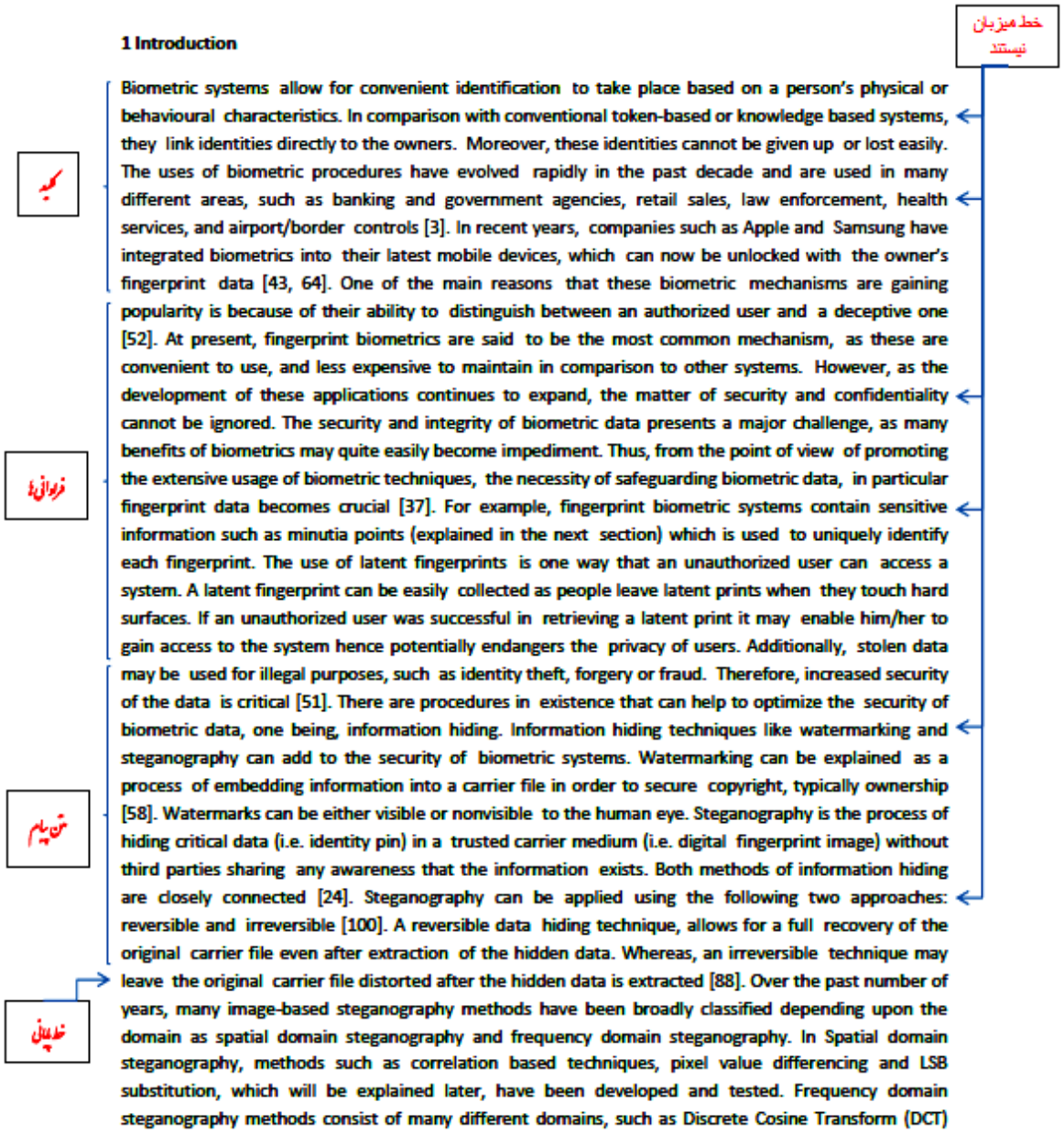
Z =
100110111111000111010010001000000110111100101
010010111

حال رشته حاصل‌شده به بلوک‌های به‌طول شش افراز می‌شود و نظیر هرکدام سه‌تایی موردنظر مشخص می‌گردد، که به‌ترتیب (از چپ به راست) عبارت‌اند از:

(2, 1, 2), (3, 3, 3), (0, 1, 3), (1,0, 2), (0, 2, 0),

(0, 1, 2), (3, 3, 0), (2, 2, 2), (1, 1, 3)

حال یک متن دلخواه برای متن پوششی انتخاب می‌شود [۱۱]. پس از درج‌های مزبور، متن پوششی که در شکل (۱) آمده، حاصل می‌شود.



شکل ۱. متن پوششی حاوی پیام.

۴-۲- نسبت ظرفیت این الگوریتم

حال به بررسی نسبت ظرفیت این الگوریتم پرداخته می‌شود. نسبت ظرفیت درج اطلاعات یک الگوریتم پنهان‌نگاری از رابطه زیر به‌دست می‌آید [۵].

$$\text{نسبت ظرفیت} = \frac{\text{تعداد بایت‌های پنهان شده}}{\text{تعداد کل بایت‌های متن پوششی}}$$

برای اندازه‌گیری نسبت ظرفیت این الگوریتم ۲۰۰ متن مختلف انتخاب شدند. با بررسی این متون دیده شد که تقریباً ۶۹٪ خطوط یک متن خط میزبان (HL) هستند. همچنین در هر سطر از این متون تقریباً ۸۰ تا ۹۰ کاراکتر وجود داشت پس در هر سطر به‌طور متوسط ۸۵ کاراکتر وجود دارد [۱۰]. همچنین توجه شود که با روش ارائه شده در هر سطر ۶ بیت یعنی $\frac{۶}{۸} = ۰/۷۵$ بایت ذخیره می‌شود؛ بنابراین نسبت ظرفیت صرفاً روش درج برابر با رابطه زیر می‌باشد:

$$\frac{1}{164} \cong \frac{69}{100} \times \frac{0/75}{85}$$

توجه شود که به‌عنوان مثال نسبت ظرفیت درج اطلاعات در الگوریتم ارائه شده توسط لیو برابر $\frac{۱}{۷۵۴}$ است [۳]. برای به‌دست آوردن ظرفیت درج الگوریتم پنهان‌نگاری ارائه شده متن‌های

متعددی انتخاب شدند. به‌عنوان مثال طول رشته دودویی نظیر یکی از متن‌ها ۵۹۲۰ بود. طول کد هافمن نظیر آن ۳۰۶۴ شد. برای درج کد هافمن نظیر آن به ۵۱۱ خط، برای درج فراوانی کد هافمن به ۴۴ خط و برای درج کلید الگوریتم رمزگذاری به ۶ خط (مجموعاً ۵۶۱ خط) نیاز بود. بنابراین نسبت ظرفیت درج الگوریتم ارائه شده ۱/۵۵٪ است. زیرا

$$\frac{5920}{561 \times 85 \times 8} \cong 1/55\%$$

با انتخاب متن‌های بزرگ‌تر نسبت ظرفیت درج بهبود می‌یابد. به‌عنوان‌مثال وقتی طول رشته متن پیام ۱۱۸۴۸ بود، طول کد هافمن نظیر آن ۶۱۳۱ شد. برای درج این کد به ۱۰۲۲ خط به‌علاوه ۴۴ خط برای درج فراوانی‌های کد هافمن و ۶ خط برای درج کلید رمز (در مجموع ۱۰۷۲ خط) نیاز است. در نتیجه نسبت ظرفیت درج آن ۱/۶۲٪ به‌دست آمد.

$$\frac{11848}{1072 \times 85 \times 8} \cong 1/62\%$$

بنابراین به‌طور متوسط نسبت ظرفیت درج در حدود ۱/۵۵٪ حاصل شد.

۴-۳- تجزیه‌وتحلیل و امنیت الگوریتم

استفاده از متن‌ها برای پنهان‌سازی اطلاعات در آن‌ها یکی از قدیمی‌ترین روش‌های پنهان‌نگاری اطلاعات است. روزانه در حدود ۱۴۵ میلیارد پیام متنی در جهان از طرق مختلف از قبیل رایانامه‌ها، پیامک‌ها، نرم‌افزارهای پیام‌رسان، و ... جابه‌جا می‌شوند. استفاده روزمره و گسترده از متن‌ها در ارتباطات باعث می‌شود که متن‌ها و پرونجاهای متنی به‌عنوان یکی از مهم‌ترین داده‌ها برای پنهان کردن اطلاعات در آنان مورد استفاده قرار بگیرند. به‌خصوص ارسال متن‌ها به‌صورت چاپ‌شده کمترین سوءظن را ایجاد می‌کند. گرچه پنهان‌نگاری متنی از سایر روش‌های پنهان‌نگاری قدیمی‌تر است اما از آنجایی‌که تغییر در متن یا پَروَنجاهای متنی به‌راحتی مشخص می‌شود، پنهان‌نگاری در متن به یکی از مشکل‌ترین پنهان‌نگاری‌ها تبدیل‌شده است.

اکثریت روش‌های ارائه شده برای پنهان‌نگاری اطلاعات صرفاً قابل استفاده برای متن‌های الکترونیکی‌اند و درصورتی‌که متن چاپ شود نمی‌توان اطلاعات مخفی را از درون‌متن پوششی استخراج کرد. الگوریتم ارائه شده در این مقاله این امکان را دارد که در غالب پرونجای PDF یا به‌صورت چاپ‌شده مورد استفاده قرار گیرد.

عملکرد یک الگوریتم پنهان‌نگاری اطلاعات در متن با مؤلفه‌های مختلفی ازجمله عدم تشخیص آن با چشم غیرمسلح، قدرت در برابر دست‌کاری، امنیت اطلاعات و ظرفیت پنهان‌سازی اندازه‌گیری می‌شود [۱۲].

برای بررسی کارایی یک الگوریتم پنهان‌نگاری، حملات مختلف نهان‌کاوی بر روی متنی که اطلاعات در آن پنهان‌شده است، انجام می‌شود. ازجمله این حملات می‌توان به حمله بصری، حمله ساختاری، حمله آماری، حمله تایپ دوباره و ... اشاره کرد. در ادامه مقاومت الگوریتم ارائه شده در این مقاله در برابر حملات فوق، مورد بررسی قرار خواهد گرفت.

یکی از مهم‌ترین عوامل در پنهان‌نگاری این است که پَروَنجای اصلی پوششی و پَروَنجای آلوده شده از نظر ظاهری بسیار به یکدیگر شبیه باشند. در یک متن تراز شده فاصله‌هایی وجود دارند که از نظر اندازه مقداری از بقیه فاصله‌ها بزرگ‌تراند تا طول خطوط با دیگر خطوط برابر شود (فاصله‌های اضافه‌شده که در این مقاله با AS نمایش داده شده‌اند). جای این فاصله‌های اضافه شده بسته به نرم‌افزارهای مختلف، متفاوت است [۱۰]. بنابراین تغییر مکان آنان اختلاف بسیار ناچیزی در متن ایجاد می‌کند و لذا متن آلوده شده با متن اصلی شباهت زیادی خواهد داشت. این امر را می‌توان به کمک روش‌های اندازه‌گیری شباهت بررسی کرد. در صورت استفاده از روش‌های اندازه‌گیری شباهت ازجمله روش نیدلمن وانش[1] یا اسمیت- واترمن[2] ملاحظه می‌شود، که متن اصلی و متنی که اطلاعات در درون آن مخفی شده‌اند از شباهت خیلی زیادی برخوردارند. بنابراین جابه‌جا شدن فاصله‌های اضافه شده در متن، شکی در نهان‌کاوان ایجاد نمی‌کند و لذا در برابر حمله نهان‌کاوی بصری مقاوم خواهد بود. با توجه به این‌که مشخص نیست که فاصله‌های اضافه شده در کجای خط هستند و با توجه به این‌که تغییری در ویژگی‌های آماری متن اعمال نمی‌شود و به‌طورکلی تعداد فراوانی کلمات، تعداد و مکان فاصله‌ها (معمولی)، مکان خطوط و غیره تغییری نمی‌کنند این نتیجه حاصل می‌شود که روش ارائه شده، در برابر حملات آماری نیز مقاوم است. روش پیشنهادی هیچ کلمه یا بخشی از متن را بر اساس زبان‌شناسی تغییر نمی‌دهد یا چیزی را جایگزین آن نمی‌کند، همچنین به‌جای کلمه اصلی از عبارت اختصاری استفاده نمی‌کند، بنابراین ساختار متن پوششی بدون تغییر باقی خواهد ماند. این امر باعث می‌شود که این روش در برابر حملات ساختاری نیز مقاومت داشته باشد. یکی دیگر از روش‌های نهان‌کاوی، حمله دوباره تایپ کردن متون است. اگر برای پنهان‌نگاری از اضافه کردن کاراکترها استفاده شود، حجم پَروَنجای آن بیشتر خواهد شد. درصورتی‌که این متن دوباره تایپ شود، چون حجم پَروَنجای آن اختلاف چشمگیری دارد، شک نهان‌کاوان برانگیخته خواهد شد و لذا به وجود پیام پی می‌برند. در

[1] Needlemen-Wunch
[2] Smith-Waterman

روش ارائه شده حجم پَروَنجا اصلاً تغییر نمی‌کند، بنابراین شکی ایجاد نمی‌شود. از طرف دیگر با حمله تایپ دوباره، متن پنهان‌شده در متن از بین می‌رود ولی شکی در آلوده بودن متن ایجاد نمی‌کند و لذا باعث بسته شدن کانال ارتباطی نخواهد شد.

استفاده از الگوریتم رمزگذاری در ابتدای کار در حفظ اطلاعات محرمانه اثرگذار خواهد بود. درنهایت از ویژگی‌های حائز اهمیت این روش می‌توان به قابلیت چاپی بودن آن اشاره کرد. مسئله حائز اهمیت دیگری که باید مورد بررسی قرار گیرد نسبت ظرفیت درج است که در بخش قبل به آن اشاره شد. در ادامه مقایسه‌ای با برخی الگوریتم‌های موجود خواهیم داشت تا نشان دهیم ظرفیت درج الگوریتم ارائه شده نسبت به آنان بیشتر خواهد بود. برخی روش‌های پنهان‌نگاری از نسبت ظرفیت درج پایینی نسبت به روش ارائه شده برخوردارند [۲]. روش‌های پنهان‌نگاری به دو دسته تقسیم می‌شوند یا همانند روش ارائه شده‌اند که علائمی را در متن اضافه یا، جابه‌جا می‌کنند یا روش‌هایی هستند که از قواعد زبان‌شناسی استفاده می‌کنند. مثلاً از مترادف یک کلمه برای درج استفاده می‌کنند، دو کلمه به‌عنوان‌مثال قدرت و نیرو که مترادف‌اند را در نظر گرفته و در هر جای متن که بخواهیم ۱ را درج کنیم از نیرو و در هر جا که بخواهیم صفر را درج کنیم از کلمه قدرت استفاده می‌کنیم. این روش‌ها را روش‌های پنهان‌نگاری زبان‌شناسی می‌نامند. ظرفیت درج اکثر الگوریتم‌های پنهان‌نگاری زبان‌شناسی نسبت به روش ارائه شده پایین‌تر است ولی در اینجا به مقایسه روش ارائه شده با مقالات مشابه پرداخته می‌شود. در جدول (۲) مقایسه روش ارائه شده در این مقاله با چند روش دیگر ارائه شده است.

برای مقایسه روش‌های موجود در پنهان‌نگاری یک صفحه از یک کتاب، در نظر گرفته می‌شود که دارای ۳۰ سطر در هر صفحه باشد و به‌طور متوسط هر خط آن ۸۵ کاراکتر داشته باشد. حال روش‌های مختلفی برای پنهان‌نگاری، در نظر گرفته و به بررسی نسبت ظرفیت درج آن‌ها پرداخته می‌شود و با نسبت ظرفیت درج الگوریتم ارائه شده، مقایسه می‌گردد. در مقاله [۱۳] که با جابه‌جا کردن سطرها به سمت بالا یا پایین اعداد صفر و یک درج می‌شوند، می‌توان در این صفحه ۳۰ بیت (۰ یا ۱) درج شود و لذا نسبت درج آن در حدود ٪۰/۱۵ $\cong \frac{۳۰}{۸\times۳۰\times۸۵}$ است.

در مقاله [۱۰]، نسبت ظرفیت درج بدون استفاده از فشرده‌سازی ٪۰/۴ است و در صورت استفاده از فشرده‌سازی ٪۱/۰۷ خواهد بود. نسبت ظرفیت درج الگوریتم ارائه شده برابر ٪۱/۵۵ هست. مابقی نسبت ظرفیت درج الگوریتم‌های جدول (۲) از مقاله ماهاتو استخراج شده است [۱۴].

جدول ۲. جدول مقایسه نسبت ظرفیت درج برخی روش‌های پیشین در پنهان‌نگاری [۶].

ردیف	نام روش	نسبت ظرفیت درج
۱	Bolshakov [۶]	٪۰/۰۳
۲	Winstein [۱۵]	٪۰/۱۵
۳	Liu et al [۳]	٪۰/۱۳
۴	Mahato et al [۱۴]	٪۰/۲۲
۵	Brassil et al [۱۳]	٪۰/۱۵
۶	Khosravi et al [۱۰]	٪۱/۰۷
۷	روش ارائه شده	٪۱/۵۵

۵- نتیجه‌گیری

در این مقاله روش جدیدی برای پنهان‌نگاری در متون ارائه گردید که با استفاده از تغییر مکان فاصله‌های اضافه شده در یک خط، اطلاعات در آن مخفی می‌گردد. از آنجایی‌که نرم‌افزارهای مختلف تایپ به شکل‌های مختلفی این فاصله‌ها را در متن جایگذاری می‌کنند، تغییر مکان این فاصله‌ها شکی ایجاد نمی‌کند. از ویژگی‌های این روش مقاومت آن در برابر برخی از حملات نهان‌کاوی است.

برخی از الگوریتم‌های پنهان‌نگاری برای درج، یک متن پوششی ایجاد می‌کنند و لذا نمی‌توانند از هر متنی برای درج استفاده کنند. در حالی‌که روش ارائه شده در این مقاله می‌تواند از هر متنی برای درج استفاده کند.

نسبت ضریب درج اطلاعات در این روش از برخی روش‌های ارائه شده بیشتر است. هیچ محدودیتی در اندازه متن پیام محرمانه در آن وجود ندارد. قابلیت استفاده به‌صورت چاپی و استفاده برای هر زبانی از دیگر مزایای این روش است.

۶- مرجع‌ها

[1] Krenn, J. R. “Steganography and Steganalysis”; http://www.Krenn.nl/univ/cry/steg/article.pdf, 2004.

[2] Xiang, L.; Wu, W.; Li, X.; Yang, C. “A Linguistic Steganography Based on Word Indexing Compression and Candidate Selection”; Multimed. Tools Appl. 2018, 77, 28969–28989.

[3] Liu, T. Y.; Tsai, W. H. “A New Steganographic Method for Data Hiding in Microsoft Word Documents by a Change Tracking Technique”; IEEE Trans. Inf. Foren. Sec. 2007, 2, 24–30.

[4] Yang, W. C.; Chen, L. H. “A Steganographic Method via Various Animations in PowerPoint Files”; Multimed. Tools Appl. 2015, 74, 1003–1019.

[5] Khairullah, M. “A Novel Text Steganography System Using Font Color of the Invisible Characters in Microsoft Word Documents”; Int. C. Comp. Elec. Eng. 2009, 482-484.

[6] Bolshakov I. A. “A Method of Linguistic Steganography Based on Collocational-Verified Synonymy”; Fridrich J. (eds) Information Hiding. IH 2004. Lecture Notes in Computer Science, vol 3200. Springer, Berlin, Heidelberg 2004.

[7] Shahreza, M. H. S.; Shahreza, M. S. “A New Approach to Persian/Arabic Text Steganography”; Proc. 5th IEEE/ACIS Int. Conf. Comput. Inform. Sci. and 1st IEEE/ACIS Int. Workshop on Component-Based Software Engineering, Software Architecture and Reuse, 2006, 310-315.

[8] Shahreza, M. H. S.; Shahreza, M. S. “A New Synonym Text Steganography”; Int. Conf. Intelligent Information Hiding and Multimedia Signal Processing, 2006, 1524-1526.

[9] Lee, I. S.; Tsai, W. H. “A New Approach to Covert Communication *via* PDF Files”; Signal Process 2010, 90, 557–565.

[10] Khosravi, B.; Nazarkardeh, Kh. “A New Method for Pdf Steganography in Justified Texts”; J. Inf. Secur. Appl. 2019, 45, 2, 61-70.

[11] Douglas, M.; Bailey, K.; Leeney, M.; Curran, K. “An Overview of Steganography Techniques Applied to the Protection of Biometric Data”; Multimed. Tools Appl. 2018, 77, 13, 17333-17373.

[12] Maji, G.; Mandal, Sh. “A Forward email Based High Capacity Text Steganography Technique Using a Randomized and Indexed Word Dictionary”; Multimed. Tools Appl. 2020, 79, 26549–26569.

[13] Brassil, J. T.; Low, S.; Maxemchuk, N. F.; O'Gorman, L., "Electronic Marking and Identification Techniques to Discourage Document Copying”; IEEE J. Selected Areas Commun. 1995, 13, 1495-1504.

[14] Mahato, S.; Khan, D. A.; Yadav, D. K. “A Modified Approach to Data Hiding in Microsoft Word Documents by Change-Tracking Technique”; J. Comput. Inf. Sci. 2020, 32, 216-224.

[15] Winstein, K. “Lexical Steganography Through Adaptive Modulation of the Word Choice Hash”; Secondary education at the Illinois Mathematics and Science Academy, 1999, http://alumni.imsa.edu/~keithw/tlex/lsteg.ps. Accessed 20 March 2017.

[16] Kumar R.; Singh, H. “Recent Trends in Text Steganography with Experimental Study”; Gupta B., Perez G., Agrawal D., Gupta D. (eds) Handbook of Computer Networks and Cyber Security. Springer, Cham, 2020.

[17] Shamalizadeh Baei, M. A.; Norozi, Z.; Sabzinezhad, M.; Karami M. R. “Designing an Image Steganography Algorithm Based on Entropy and ELSB2”; Adv. Defence Sci. & Technol. 2018, 02, 39-50 (In Persian).