فصلنامه پژوهش‌های حفاظتی ـ امنیتی دانشگاه جامع امام حسین(علیه‌السلام)
سال پنجم، شماره‌ی ۱۸ (تابستان ۱۳۹۵): صص ۴۳ – ۵۹

ارائه یک روش جدید جهت پنهان‌نگاری اطلاعات در فایل متنی فارسی

طاهره صفدری[۱]
علی‌محمد لطیف[۲]

تاریخ دریافت: ۱۳۹۴/۱۲/۱۵
تاریخ پذیرش: ۱۳۹۵/۰۵/۱۵

چکیده

در این مقاله یک روش جدید برای پنهان‌نگاری اطلاعات در فایل متنی فارسی ارائه‌شده است. در این روش با ایجاد تغییراتی در فاصله نقاط حروف دونقطه‌ای «ت، ق، ی» و حروف سه‌نقطه‌ای «ژ، ش، ث، پ، چ» به پنهان‌نگاری اطلاعات در متن فارسی پرداخته می‌شود. در هنگام درج، اگر بیت صفر باشد، هیچ تغییری در حرف موردنظر ایجاد نمی‌شود و اگر بیت یک باشد، فاصله بین نقاط حرف موردنظر کم‌تر می‌شود. در پایان جهت مصونیت فایل متنی در برابر حمله‌های تغییر فونت و اندازه فونت، فایل خروجی به‌صورت تصویر ذخیره می‌گردد. در مرحله استخراج با عملگرهای ریخت‌شناسی تصویری، اطلاعات درج‌شده استخراج می‌گردد. درصد پنهان‌سازی اطلاعات در این روش ۱۷/۰۴ است که نسبت به سایر روش‌های موجود بهبود داده شده است.

کلید واژه‌ها: *پنهان‌نگاری، متن فارسی، مخفی‌سازی داده‌ها*

۱- کارشناسی ارشد فناوری اطلاعات، گروه مهندسی کامپیوتر، دانشگاه یزد. taherehsafdari@stu.yazd.ac.ir
۲- استادیار گروه مهندسی کامپیوتر، دانشگاه یزد. alatif@yazd.ac.ir

مقدمه

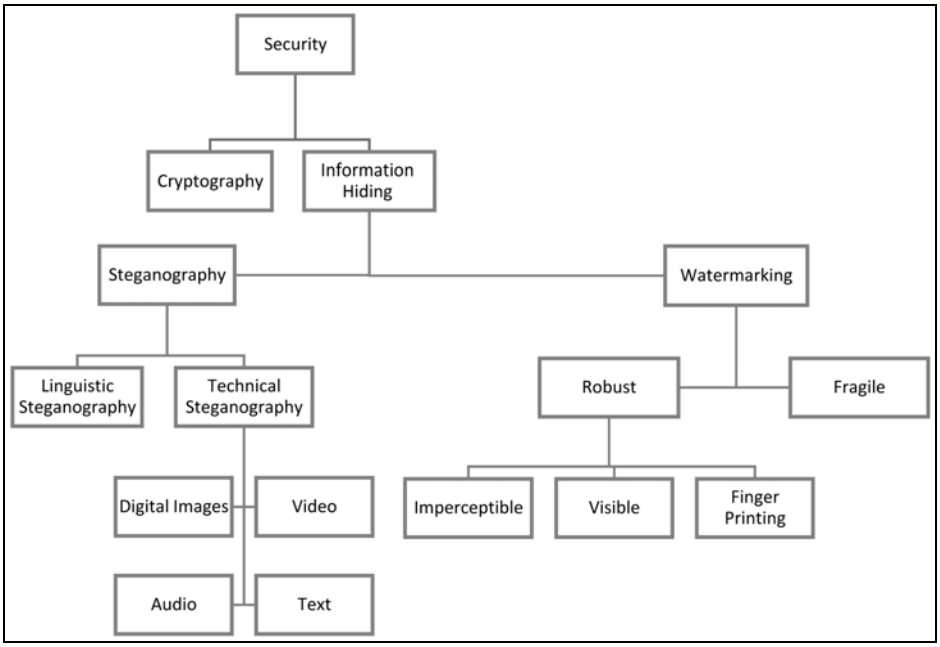
پیشرفت سریع علوم کامپیوتر، شبکه‌های کامپیوتری، ارتباطات و گسترش روزافزون کاربرد سیستم‌های چندرسانه‌ای دیجیتالی باعث ایجاد تغییرهای فراوانی در زندگی بشر شده است و امروزه اینترنت به محبوب‌ترین کانال برای تبادل اطلاعات تبدیل شده است. توسعه ابزارهای نرم‌افزاری و سخت‌افزاری علاوه بر داشتن مزایا و استفاده‌های فراوان مشکل‌هایی را برای امنیت داده ایجاد کرده است.

علاوه بر رمزنگاری، پنهان‌نگاری[1] داده‌ها نقش مهمی در تضمین امنیت پیام‌های انتقالی ایفا می‌کند. استگانوگرافی از دو واژه یونانی Steganos به معنی پوشیده و Graptos به معنی نوشتن گرفته شده است (Mohamed, 2014, 79-87) و اصطلاح استگانوگرافی به معنای پنهان کردن اطلاعات در بین اطلاعات دیگر می‌باشد. پنهان‌نگاری دارای روش‌های گسترده‌ای برای مخفی کردن اطلاعات در رسانه‌های مختلف می‌باشد. از پنهان‌نگاری در متن، تصویر، صوت، سیگنال‌ها و رسانه‌های دیگر استفاده می‌شود.

طبقه‌بندی‌های مختلفی برای پنهان‌نگاری داده‌ها وجود دارد که این طبقه‌بندی‌ها با توجه به هدف و کاربرد انواع پنهان‌نگاری متفاوت می‌باشد. یکی از طبقه‌بندی‌ها که توسط چداد[2] و همکاران او در سال ۲۰۰۹ میلادی بر اساس امنیت[3] سیستم ارائه‌شده است در شکل ۱ مشاهده می‌گردد (Cheddad, 2009, 2465-2478).

بر اساس این طبقه‌بندی امنیت به دو زیرمجموعه رمزنگاری[4] و پنهان کردن اطلاعات[5] تقسیم می‌شود. پنهان کردن اطلاعات خود به دو شاخه واترمارکینگ[6] و پنهان‌نگاری تقسیم می‌شود. پنهان‌نگاری دارای دو روش زبانی[7] و فنی[8] است که در روش‌های فنی می‌توان از انواع رسانه مانند ویدئو، صوت، تصاویر دیجیتال و متن استفاده کرد. واترمارکینگ به دو شاخه شکننده[9] و مقاوم[10] تقسیم می‌شود. واترمارکینگ قوی می‌تواند از نوع نامحسوس[11]، مرئی[12] و انگشت‌نگاری[13] باشد.

1. Steganography
2. cheddad
3. Security
4. Cryptography
5. Information Hiding
6. Watermarking
7. Linguistic Steganography
8. Technical Steganography
9. Fragile
10. Robust
11. Imperceptible
12. Visible
13. Finger printing



شکل ۱ ـ طبقه‌بندی امنیت (Cheddad, 2009)

بیان مسئله

در این مقاله یک روش جهت درج و استخراج اطلاعات در متن فارسی ارائه‌شده است. شکل برخی از حروف موجود در متن با توجه به اطلاعات قابل درج تغییر داده می‌شود. پس از درج برای حفاظت از حمله‌های تغییر فونت و اندازه فونت، این فایل به تصویر تبدیل می‌شود و ارسال می‌گردد. لازم به ذکر است عملیات استخراج اطلاعات در این روش با استفاده از عملیات پردازش تصویر صورت می‌گیرد.

اهمیت و ضرورت تحقیق

پنهان‌نگاری برخلاف رمزنگاری که فایل حفاظت‌شده را حساس جلوه می‌دهد و جلب‌توجه می‌کند، از ناآگاهی افراد برای جلوگیری از دستیابی آن‌ها به اطلاعات خاص بهره می‌برد. تفاوت اصلی رمزنگاری و پنهان‌نگاری آن است که در رمزنگاری هدف مخفی کردن محتویات پیام است و نه به‌طورکلی وجود پیام؛ اما در پنهان‌نگاری هدف مخفی کردن هرگونه نشانه‌ای از وجود پیام است. به‌عنوان‌مثال اگر شخصی به متن رمزنگاری‌شده‌ای دسترسی پیدا کند، به‌هرحال متوجه می‌شود که این متن حاوی پیام رمزی می‌باشد؛

اما در پنهان‌نگاری شخص از وجود پیام مخفی در متن اطلاعی حاصل نمی‌کند. بدیهی است در موارد حساس ابتدا متن رمزنگاری می‌شود و سپس در متن پنهان می‌گردد.

موارد مهم در طراحی یک روش پنهان نگاری

با توجه به نوع کاربرد، چندین ویژگی برای بیان کارایی روش‌های پنهان‌نگاری در نظر گرفته می‌شود که البته این ویژگی‌ها در کاربردهای مختلف، متفاوت می‌باشند. در این بخش به معرفی چندین ویژگی مهم و مشترک روش‌های پنهان‌نگاری پرداخته می‌شود.

۱- شفافیت[1]: در عمل پنهان نگاری، لازم است تا اطلاعات به‌گونه‌ای در رسانه پوششی درج شوند که اطلاعات درج‌شده از لحاظ شهودی قابل‌تشخیص و ادراک نباشد. میزان شباهت رسانه پوششی قبل و بعد از درج اطلاعات، شفافیت نامیده می‌شود (Tao, 2004, 133-144).

۲- ظرفیت[2]: حجم اطلاعات قابل درج در رسانه پوششی را ظرفیت گویند. هر روشی که ارائه می‌شود باید به‌گونه‌ای باشد که حجم قابل قبولی از اطلاعات را در رسانه پوششی پنهان کند (Siraj, 2015, 3755–3760).

۳- مقاومت[3]: مقاومت یک سیستم پنهان‌نگاری به معنای این است که پیام پنهان‌شده در مقابل اعمال تغییرات ناخواسته و غیرعمدی که وجود نویز در طول مسیر انتقال به وجود می‌آورد و یا اعمال تغییرات عمدی که توسط حمله‌کننده فعال به‌منظور تغییر پیام یا از بین بردن آن انجام می‌گیرد مقاومت لازم را داشته باشد (Siraj, 2015, 3755–3760).

سؤال‌های تحقیق

۱- آیا با ایجاد تغییرات در طراحی فونت فارسی، می‌توان پیام‌ها را در متن مخفی کرد؟

۲- آیا می‌توان روش جدیدی ارائه کرد که بدون حساسیت به قلم خاصی داده‌ها را پنهان کرد؟

۳- کارایی این روش، نسبت به روش‌های موجود چگونه است؟

1. Transparency
2. Capacity
3. Robustness

پیشینه تحقیق

در مقایسه با دیگر رسانه‌ها مانند صوت، تصویر و ویدئو فعالیت‌های کم‌تری برای پنهان‌نگاری اطلاعات در متن صورت گرفته است. در میان روش‌های موجود، تعداد کمی از آن‌ها روی متون فارسی مورداستفاده قرار می‌گیرند. در ادامه به برخی از روش‌های پنهان‌نگاری در متن فارسی اشاره مختصری خواهد شد.

۱- استفاده از برخی کاراکترهای خاص در متن

در این‌گونه روش‌ها برای پنهان‌نگاری اطلاعات، از برخی کاراکترهای خاص موجود در متن، مانند حرف اول هر پاراگراف استفاده می‌شود. این روش‌ها ظرفیت محدودی دارند و نمی‌توان حجم زیادی از اطلاعات را در متن پنهان کرد (Garg, 2011, 129-138).

۲- شیفت کلمات

در سال ۲۰۰۳ کیم[1] و همکارانش روشی ارائه کردند که با استفاده از شیفت کلمات به پنهان‌نگاری اطلاعات در متن می‌پرداخت. در این روش ابتدا کلمات متن استخراج می‌شود و سپس هر کلمه با توجه به بیت ورودی، به سمت راست یا چپ شیفت داده می‌شود. عیب این روش این است که تغییرات زیادی در متن ایجاد می‌کند. بنابراین امنیت آن به شدّت کاهش می‌یابد. هم‌چنین این روش مقاومت کمی نسبت به تغییرات دارد (Kim, 2003, 775-779).

۳- شیفت خطوط

در سال ۲۰۰۴ آل عطار[2] روشی برای پنهان‌نگاری متن با استفاده از شیفت خطوط ارائه کرد. به‌این‌ترتیب که ابتدا تمامی خطوط استخراج می‌شود و از هر سه خط متوالی، خط‌های بالا و پایین به‌عنوان خطوط کنترلی و خط میانی به‌عنوان خط حامل در نظر گرفته می‌شود. به‌عبارت‌دیگر در این روش، خط میانی با توجه به مقدار بیتی که قرار است پنهان شود به‌اندازه ۱/۳ اینچ به سمت بالا یا پایین شیفت داده می‌شود. عیب این روش همان خطوط کنترلی هستند که باعث کاهش ظرفیت داده‌های پنهان‌شده می‌شوند؛ زیرا در این روش به ازای هر سه خط متوالی، تنها خط میانی برای ذخیره‌سازی یک بیت استفاده

1. Kim
2. Alattar

می‌شود. بدیهی است اگرچه در این روش شیفت این میزان اندک از خطوط بـا چشـم دیـده نمـی‌شـود؛ امـا توسط اسکنرهای خاص قابل‌تشخیص است (Alattar, 2004, 685-695).

۴- روش نقطه‌ها

در سال ۲۰۰۶ شیرعلی روش نقطه‌ها را ارائه کرد. در این روش اطلاعاتی که بایـد مخفـی شـوند ابتـدا فشرده می‌شوند؛ سپس در متن موردنظر اولین حرف نقطه‌دار پیدا می‌شـود. اگـر بیـت قابـل درج صـفر بـود کاراکتر بدون تغییر باقی می‌ماند. اما اگر بیت قابل درج یک بود نقطه مربوط به کاراکتر کمی به سـمت بـالا جابه‌جا می‌شود. در پایان برای منحرف کردن توجه خوانندگان بعد از مخفـی کـردن همـه اطلاعـات، نقـاط مربوط به بقیه کاراکترها به‌صورت تصادفی تغییر می‌کنند. برای کاراکترهایی که دو یا سه نقطه دارند، همـه نقاط تغییر مکان می‌دهند (Shirali-Shahreza, 2006, 1692-1698). از معایـب ایـن روش مـی‌تـوان از بـین رفتن اطلاعات در نگارش مجدد و عدم مقاومت روش در برابر حمله تغییر فونت را نام برد.

۵- تغییر شکل ظاهری برخی از کاراکترها

در سال ۱۳۸۷ داورزنی و همکارش روش جدیدی را ارائه کردند. در این روش که روی تصاویر متن کار می‌کند از چهار حرف «ر، ز، ژ، و» استفاده می‌شود. این حروف دارای شیب خاصی هستند و می‌توان از ایـن شیب برای پنهان‌نگاری داده‌ها استفاده کرد (داورزنی، ۱۳۸۷، ۸-۱).

سه حرف «ر، ز، ژ» تنها در تعداد نقاط با یکدیگر متفاوت هستند. بنابراین در ایـن روش پـارامتری کـه برای شیب این سه حرف در نظر گرفته می‌شود از پارامتر شیب مربوط به حرف «و» جدا مـی‌شـود. در ایـن روش، ابتدا حروف شیب‌دار مذکور از تصویر متن اصلی استخراج می‌شود. اگر بیت قابل درج صفر بود، شیب حرف را تغییر نمی‌دهد؛ ولی اگر بیت قابل درج یک بود شیب حرف را تغییر می‌دهد.

۶- استفاده از روش تغییر مکان حروف نسبت به خط زمینه

در سال ۱۳۹۱ شهیدیان و همکارانشان روش جدیدی را برای پنهان‌نگاری اطلاعات در متـون فارسـی ارائه کردند. این روش از چهار حرف «ج، چ، ح، خ» استفاده می‌کند. در این روش اگر بیـت پیـام یـک بـود، موقعیت این حروف نسبت به خط زمینه تغییر پیدا می‌کند و اگر صفر بود، موقعیت آن تغییری پیدا نمـی‌کنـد (شهیدیان، ۱۳۹۱، ۱۷۹).

روش پیشنهادی

مزیت استفاده از متن به‌عنوان رسانه پوششی این است که فایل متنی فضای ذخیره‌سازی کم‌تری مصرف می‌کند و برای انتقال آن پهنای باند کم‌تری موردنیاز است (Shivania, 2015, 1401-1410). هدف اصلی این مقاله ارائه یک روش جدید برای پنهان‌نگاری اطلاعات در متون فارسی است. در ادامه پس از شرح روش پیشنهادی، مراحل درج و استخراج اطلاعات شرح داده می‌شود.

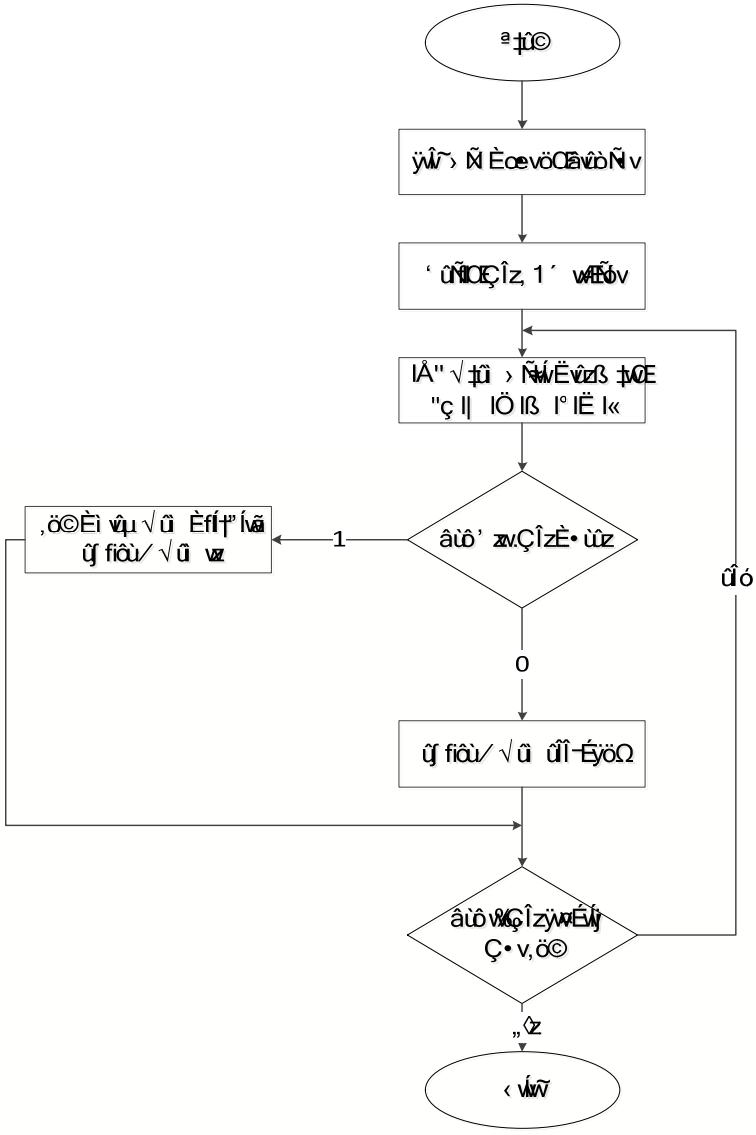
روش پیشنهادی به این صورت است که با ایجاد تغییراتی جزئی در فاصله نقاط حروف دونقطه‌ای مثل «ت، ی، ق» و حروف سه‌نقطه‌ای مثل «ژ، ش، ث، پ، چ» به پنهان‌نگاری اطلاعات در فایل متنی فارسی پرداخته می‌شود. برای این کار یک فونت جدید طراحی‌شده است؛ این فونت مشابه فونت "Tahoma" است. با این تفاوت که علاوه بر حروف موجود در این فونت، چند حرف اضافه‌تر دارد که درواقع همان حروف دونقطه‌ای و سه‌نقطه‌ای فارسی هستند که فاصله دو نقطه موجود در آن کم‌تر از حد معمول شده است. البته این تفاوت در فاصله نقاط جزئی است و جلب‌توجه نمی‌کند. طراحی این فونت جدید با استفاده از نرم‌افزار ایجاد فونت FontCreator انجام گرفته است.

درج اطلاعات

الگوریتم درج اطلاعات در این روش با استفاده از زبان برنامه‌نویسی c# پیاده‌سازی شده است؛ برنامه نوشته شده دارای دو فایل متنی است. یک فال متنی که حاوی پیام مخفی که باید در فایل متنی دوم پنهان شود. در این روش ابتدا کد اسکی متن پیام مخفی به‌صورت یک رشته بیتی استخراج می‌شود. هم‌چنین ۱۶ بیت در ابتدای این فایل، به‌عنوان بیت کنترلی برای نگه‌داشتن طول پیام مخفی مورداستفاده قرار می‌گیرد. حال فایل متنی فارسی که باید پیغام مخفی در آن پنهان شود را کاوش کرده، تا حروف موردنظر (ت، ق، ی، ژ، ش، ث، پ، چ) یافت شود. برای این‌که مشخص شود حرف موردنظر باید تغییر کند یا نه، فایل باینری پیغام بررسی می‌شود. اگر بیت قابل درج صفر باشد هیچ تغییری در حرف موردنظر ایجاد نمی‌شود. اما اگر یک باشد، حرفی که طراحی‌شده است و در آن فاصله نقاط کم‌تر از حد معمول است، جایگزین حرف موردنظر شده و این روند تا زمانی که پیغام مخفی به اتمام برسد، ادامه پیدا می‌کند. روند نمای عملیات درج اطلاعات را می‌توان در شکل ۲ مشاهده کرد.

به دلیل این‌که ایجاد تغییر در فایل متنی ساده و آسان است، پس از درج اطلاعات، فایل خروجی به فرمت تصویر تبدیل می‌شود. این کار برای امنیت بیش‌تر است؛ چراکه با این روش فایل حامل از اعمال

تغییرات مصون می‌ماند. در شکل ۳- الف متن اصلی مشاهده می‌شود و در شکل ۳- ب بیت‌های پیام مخفی و در شکل ۳- ج متن پنهان‌نگاری شده دیده می‌شود که درواقع حاصل ترکیب پیام اصلی و پیام مخفی می‌باشد. همان‌طور که مشاهده می‌شود در این یک پاراگراف از متن فارسی تعداد ۳۵ بیت پنهان‌شده است.



شکل ۲ ـ روند نمای درج اطلاعات در روش پیشنهادی

با توجه به گسترش فناوری در دنیای امروز و امکان انجام اکثر عملیات از راه دور، با استفاده از شبکه‌های جهانی و محلی، هم‌چنین عدم لزوم تمرکز همه داده‌ها در یک محل و نیاز به دستیابی به برخی از اطلاعات راه دور و هم چنین حفظ امنیت اطلاعات در زمان ارسال و دریافت، اهمیت نگهداری اطلاعات از دسترسی های غیرمجاز را بیش از پیش آشکار می سازد.

الف) متن اصلی

100010011000001001001000010100100001

ب) بخشی از بیت‌های پیام مخفی‌شده

با توجه به گسترش فناوری در دنیای امروز و امکان انجام اکثر عملیات از راه دور، با استفاده از شبکه‌های جهانی و محلی، هم‌چنین عدم لزوم تمرکز همه داده‌ها در یک محل و نیاز به دستیابی به برخی از اطلاعات راه دور و هم چنین حفظ امنیت اطلاعات در زمان ارسال و دریافت، اهمیت نگهداری اطلاعات از دسترسی های غیرمجاز را بیش از پیش آشکار می سازد.

ج) متن پنهان‌نگاری شده

شکل ۳ ـ مقایسه متن اصلی و متن پنهان‌نگاری شده

استخراج اطلاعات

فایل ورودی برای عملیات استخراج، فایلی با فرمت تصویر است و تمام عملیاتی که برای شناسایی و استخراج اطلاعات انجام می‌گیرد، عملیات پردازش تصویر می‌باشد. یکی از این عملیات مورفولوژی[1] می‌باشد که در ادامه مختصری از آن شرح داده می‌شود. مورفولوژی ریاضی به‌عنوان ابزاری به شمار می‌آید که برای استخراج اجزای مفید تصویر، که در ارائه و توصیف شکل‌های اصلی مناسب هستند به کار گرفته می‌شود. عملیات ساختاری، عملیاتی هستند که بر روی تصاویر باینری اعمال شده و هدف از آن ایجاد تغییر و یا تصحیح در اجزاء داخل یک تصویر باینری باشد. این عملیات اغلب یک مرحله قبل از عملیات پردازش نهایی انجام می‌شود. منظور از عملیات پردازش نهایی، عملیاتی است که در آن اطلاعاتی از تصویر استخراج می‌شود. از میان این عملیات در ادامه چهار نوع از آن شرح داده خواهد شد (Gonzalez, 1996):

1. Morphology

۱- عملیات گسترش[1]: عملیاتی است که باعث افزایش ابعاد اجزاء تصویر به‌اندازه یک یا چند پیکسل می‌گردد. در اثر این عمل ممکن است نقاطی که از یک تصویر باینری در اثر عواملی چون تأثیر نویز یا اعمال حد آستانه نامطلوب جا افتاده است، تصحیح گردند. به‌عنوان‌مثال ممکن است دو جزء از تصویر به یکدیگر متصل گردند.

۲- عملیات فرسایش[2]: عکس عملیات گسترش است. در این عملیات نقاط ناخواسته تصویر باینری حذف می‌شوند و سایر اجزا تصویر نیز به‌اندازه یک یا چند پیکسل نازک‌تر خواهند شد.

۳- عملیات گشودن[3]: در عملیات گشودن اجزایی از تصویر باینری که از یک اندازه تعیین‌شده کوچک‌تر باشند حذف می‌شوند، بدون آن‌که ابعاد سایر اجزا تغییر کند.

۴- عملیات بستن[4]: در این عملیات نواحی جا افتاده تصویر باینری بدون تغییر در ابعاد سایر اجزاء ترمیم می‌گردند.

تمام مراحل عملیات استخراج اطلاعات در روش پیشنهادی با استفاده از زبان برنامه‌نویسی MATLAB پیاده‌سازی شده است. برای استخراج اطلاعات در این روش مراحل زیر باید طی شود.

۱- در ابتدا مؤلفه‌های همبندی که مساحت آن از یک حد معینی بیش‌تر است حذف می‌شود، با این کار نقاط حروف (یک نقطه، دو نقطه و سه نقطه) موجود در تصویر باقی می‌مانند.

۲- روی تصویر عمل گسترش انجام می‌شود تا جفت نقاطی که فاصله آن‌ها کم‌تر از حد معمول است به هم بچسبند و موقعیت این نقاط به دست آید.

۳- اجرای مجدد گسترش موجب می‌شود تا جفت نقاطی که فاصله آن‌ها تغییر نکرده است و در همان حد معمول است، به هم بچسبند. بنابراین موقعیت این نقاط به دست می‌آید.

۴- پس از آن‌که موقعیت همه دونقطه‌ای‌های موجود در تصویر به دست آمد، این نقاط مرتب می‌شوند. به‌جای نقاطی که در مرحله ۲ یافت شده‌اند در فایل متنی عدد یک و به‌جای نقاطی که در مرحله ۳ یافت شده‌اند، عدد صفر قرار می‌گیرد. با این روش کلیه بیت‌های پنهان‌شده در متن فارسی استخراج می‌شوند. خروجی عملیات استخراج، یک فایل متنی است که بیت‌های پنهان‌شده در متن را نشان می‌دهد. روند نمای عملیات استخراج اطلاعات را می‌توان در شکل ۴ مشاهده کرد.

1. Dilation
2. Erosion
3. Opening
4. Closing

شکل ۴ ـ روند نمای استخراج اطلاعات در روش پیشنهادی

جمع‌بندی و نتیجه‌گیری

در این مقاله یک روش جدید برای پنهان‌نگاری اطلاعات در متن فارسی ارائه شد. در روش پیشنهادی با ایجاد تغییراتی جزئی در فاصله نقاط حروف دونقطه‌ای مثل «ت، ق، ی» و حروف سه‌نقطه‌ای مثل «ژ، ش، ث، پ، چ» به پنهان‌نگاری اطلاعات در متن فارسی پرداخته‌شده است. روش کار به این صورت است که اگر بیت قابل درج صفر باشد هیچ تغییری در حرف موردنظر ایجاد نمی‌شود و اگر آن بیت یک باشد فاصله بین نقاط حرف موردنظر کم‌تر می‌شود.

برای استخراج اطلاعات درج‌شده ابتدا مؤلفه‌های همبندی استخراج‌شده و با توجه به مساحت آن‌ها اشیای بزرگ تصویر حذف می‌گردد و فقط نقاط موجود در تصویر باقی می‌ماند. سپس عمل گسترش روی تصویر اعمال می‌شود. در این حالت جفت نقاطی که به هم نزدیک شده‌اند (فاصله آن کم‌تر شده است) توسط یک عملگر به هم چسبیده و به‌عنوان یک شیء شناسایی می‌شوند. به علت افزایش مساحت، این شیء نیز از تصویر حذف می‌گردد. این اشیای شناسایی‌شده نشان‌دهنده بیت پنهان‌شده صفر هستند. در مرحله بعد عمل گسترش جدیدی روی تصویر اعمال می‌گردد؛ در این مرحله جفت نقاط با فاصله معمول شناسایی می‌شوند که نشان‌دهنده بیت پنهان‌شده یک هستند.

در ادامه به ارزیابی روش پیشنهادی پرداخته و سه عامل مهم در ارزیابی روش‌های پنهان نگاری، یعنی شفافیت، ظرفیت و مقاومت را برای روش پیشنهادی بررسی کرده و تقابل بین این سه عامل موردبررسی قرار خواهد گرفت.

تقابل شفافیت، ظرفیت و مقاومت: به‌طورکلی در طراحی یک سیستم پنهان‌نگاری مناسب باید سعی شود که با توجه به کاربردها و نیازهای موجود، تعادلی را بین سه عامل مقاومت، شفافیت و ظرفیت برقرار کرد. البته باید به این نکته توجه داشت که این سه عامل را می‌توان به‌صورت سه رأس از یک مثلث در نظر گرفت که در تقابل با یکدیگر هستند. یعنی افزایش هرکدام منجر به کاهش عامل دیگر می‌شود. بنابراین معمولاً در سیستم‌های پنهان‌نگاری سعی می‌شود تا بین این سه عامل موازنه برقرار شود. با این وجود در سیستم‌های پنهان‌نگاری تأکید اصلی بر بررسی دو عامل ظرفیت و شفافیت است (Esra, 2012, 2385–2394).

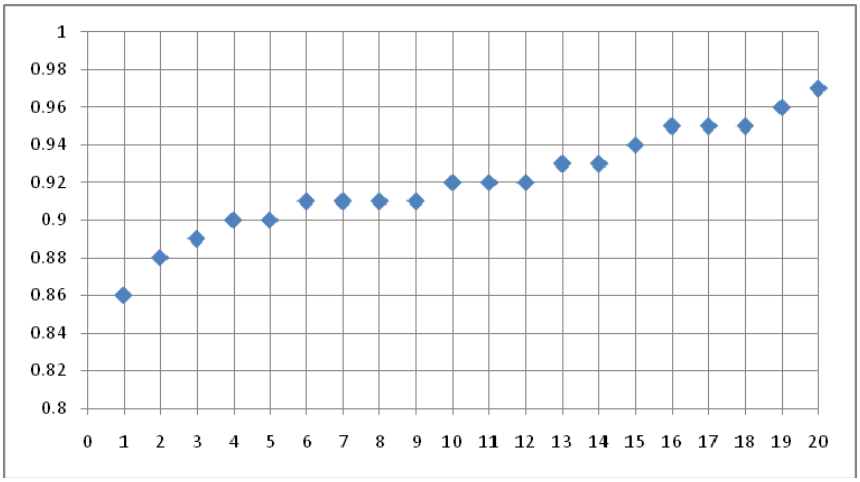
نتایج حاصل از انجام روش پیشنهادی بر روی متون فارسی نشان می‌دهد که این روش از شفافیت و ظرفیت بالایی برخوردار بوده و در مقایسه با روش‌های مشابه، حجم اطلاعات بیش‌تری را می‌توان در متن پنهان کرد. مقاومت این روش نیز در حد مطلوبی قرار دارد. در ادامه هر یک از این عوامل به‌طور جداگانه بررسی شده است.

شفافیت: برای محاسبه شفافیت روش پیشنهادی از SSIM[1] استفاده شده است. SSIM روشی جهت مقایسه تشابه دو تصویر است. برای این کار تصویر متن اصلی و متن پنهان‌نگاری شده را مقایسه کرده و با استفاده از این روش محاسبه می‌شود که تصویر متن پنهان‌نگاری شده چه میزان نسبت به تصویر متن اصلی متفاوت بوده است.

1. similarity Structural

در شکل ۵ می‌توان نمودار مقادیر SSIM را برای تصاویر بخش‌های مختلفی از متن مشاهده کرد. به‌عنوان نمونه این مقایسه بر روی ۴۰ تصویر به‌صورت دوبه‌دو (قبل و بعد از درج اطلاعات) انجام گرفته است. مقدار SSIM محاسبه‌شده برای بخش‌های مختلفی از تصویر متن متفاوت بوده است و میانگین این مقدارها در حد ۰/۹۲ محاسبه‌شده است. این میزان از تفاوت به علت ایجاد کشش در برخی از حروف رخ داده است که موجب می‌شود برخی از حروف به خط بعد منتقل شوند. البته این امر به‌ندرت اتفاق می‌افتد و تعداد کمی از خطوط متن حاوی اطلاعات مخفی با متن اصلی تفاوت دارند.

هم‌چنین نرخ درج اطلاعات در تصاویر یکسان نبوده است. یکسان نبودن نرخ درج اطلاعات در تصاویر مختلف به دلیل تعدد تکرار حروف دونقطه‌ای و سه‌نقطه‌ای در بخش‌های مختلف متن می‌باشد. تفاوت ایجادشده در فاصله نقاط بر روی این مقدار از SSIM تأثیر اندکی داشته است. هم‌چنین به دلیل این‌که فایل خروجی تصویری از متن است، ناظر انسانی قادر به تشخیص این تفاوت‌ها در متن نمی‌باشد. بنابراین می‌توان گفت شفافیت روش پیشنهادی در حدّ مطلوبی قرار دارد و این میزان از تفاوت در تصویر موجب نمی‌شود که کسی به متن پنهان‌نگاری شده مشکوک شود.



شکل ۵ ـ مقادیر SSIM برای تصویر بخش‌های مختلف متن

ظرفیت: در جدول ۱ ظرفیت روش پیشنهادی با دیگر روش‌های پنهان‌نگاری در متون فارسی مقایسه شده است. این مقایسه به‌عنوان‌مثال بر روی بخشی از روزنامه خراسان انجام گرفته است. در این جدول،

ستون ظرفیت پنهان‌سازی درواقع تعداد کاراکترهایی را که می‌توان اطلاعاتی را در آن پنهان کرد نشان می‌دهد؛ برای این کار برنامه‌ای نوشته شده است که تعداد کاراکترهایی را که در هر روش برای پنهان‌نگاری اطلاعات از آن‌ها استفاده می‌شود برمی‌شمارد. درصد پنهان‌سازی، نسبت تعداد کاراکترهایی که می‌توان اطلاعات در آن ذخیره کرد (ظرفیت پنهان‌سازی)، به تعداد کل کاراکترهای موجود در متن را نشان می‌دهد:

$$\text{درصد پنهان‌سازی اطلاعات} = \frac{\text{تعداد کاراکترهایی که می‌توان اطلاعات در آن ذخیره کرد}}{\text{تعداد کل کاراکترهای موجود در متن}}$$

جدول ۱ـ مقایسه ظرفیت در روش‌های پنهان‌نگاری برای یک نمونه متن فارسی

نام روش	تعداد کل کاراکترهای موجود در متن	ظرفیت پنهان‌سازی	درصد پنهان‌سازی
استفاده از کاراکتر خاص در متن (کلمه اول هر پاراگراف)	۳۲۳۲	۲	۰/۰۶
شیفت خطوط	۳۲۳۲	۱۷	۰/۵۲
روش نقطه‌ها	۳۲۳۲	۹۴۶	۲۹/۲۶
تغییر شکل ظاهری برخی از کاراکترها (تغییر شیب)	۳۲۳۲	۳۴۲	۱۰/۵۸
تغییر مکان حروف نسبت به خط زمینه	۳۲۳۲	۱۰۸	۳/۳۴
روش پیشنهادی	۳۲۳۲	۵۵۱	۱۷/۰۴

همان‌طور که مشاهده می‌شود درصد پنهان‌سازی اطلاعات در روش پیشنهادی ۱۷/۰۴ درصد محاسبه‌شده است که نسبت با دیگر روش‌های پنهان‌نگاری در حد بالایی قرار دارد.

مقاومت: در این روش پس از پنهان کردن اطلاعات، فایل متنی به فرمت تصویر تبدیل می‌شود. با این کار فایل حامل از تغییراتی که روی فایل‌های متنی ایجاد می‌شود (مانند تغییر اندازه، رنگ، فونت)، مصون می‌ماند. بنابراین امکان ایجاد تغییر در متن وجود ندارد. به همین دلیل می‌توان گفت این روش از مقاومت مطلوبی برخوردار است.

پیشنهادها

برای بهبود عملکرد روش و ادامه تحقیق در مورد آن، موارد زیر به‌عنوان پیشنهاد مطرح می‌گردد:

۱- این روش برای نوشته‌های دست‌نویس کاربردی نیست، چراکه احتمال خطا در این‌گونه نوشته‌ها زیاد است. به‌عنوان کارهای آتی می‌توان تغییراتی ایجاد کرد که این روش برای نوشته‌های دست‌نویس نیز مورداستفاده قرار گیرد.

۲- میتوان از آن برای عملیات استخراج اطلاعات از سایر عملگرهای تصویر استفاده کرد.

منابع و مآخذ:

داورزنی، رضا و خشایار یغمایی (۱۳۸۷)، ***واترمارکینگ متن فارسی بر اساس کدینگ کاراکتر***، در: چهاردهمین کنفرانس سالانه انجمن کامپیوتر ایران.

شهیدیان، علی اصغر و دیگران (۱۳۹۱) ***استفاده از تنوع قواعد دستوری و نوشتاری برای پنهان‌نگاری در زبان فارسی***، در: نخستین کنفرانس بین‌المللی پردازش خط و زبان فارسی.

Mohamed, A. A (2014) An Improved Algorithm for Information Hiding based on Features of Arabic Text: A Unicode Approach, in: Egyptian Informatics Journal, Vol. 15, Issue 2, pp. 79-87.

Cheddad, J. Condell, K. Curran, P. M. Kevitt (2009) A Skin Tone Detection Algorithm for an Adaptive Approach to Steganography, in: Signal Processing, vol. 89, pp. 2465-2478.

Tao, P. Eskicioglu, A. M (2004) A Robust Multiple Watermarking Scheme in the Discrete Wavelet Transform Domain, in: Proceedings of the SPIE, Vol. 5601, pp. 133-144.

Siraj, S. Sudheer, S. K. Mahadhevan Pillai. V. P (2015) Performance and Analysis of High Capacity Steganography of Color Images Involving Wavelet Transform, in: Optik-International Journal for Light and Electron Optics, Vol. 126, Issue. 23, pp 3755–3760.

Garg, M (2011) A Novel Text Steganography Technique based on HTML Documents, in: IJ Advanced Science and Technology, vol. 35, pp. 129-138.

Kim, Y. Moon, K. Oh, I (2003) A Text Watermarking Algorithm based on Word Classification and Inter word Space Statistics, in: Proceedings of the Seventh International Conference on Document Analysis and Recognition, pp.775-779.

Alattar, A.M. Alattar, O.M (2004) Watermarking Electronic Text Documents Containing Justified Paragraphs and Irregular Line Spacing, in: Proceedings of SPIE, Vol. 5306, pp. 685-695.

Shirali-Shahreza, M. H. Shirali-Shahreza, M (2006) A New Approach to Persian/Arabic Text Steganography, in: Proceedings of 5th IEEE/ACIS International Conference on Computer and Information Science, Vol. 4, Issue. 11, pp. 1692-1698.

Shivania, Yadava, V. K. Bathamb, S (2015) A Novel Approach of Bulk Data Hiding Using Text Steganography, in: Procedia Computer Science, 3rd International Conference on Recent Trends in Computing, Vol. 57, pp. 1401-1410.

C. Gonzalez, R. E. Woods, S. L. Eddins (1996) Digital Image Processing, Englewood Cliffs, NJ.

Esra, S. Hakan, I (2012) A Compression-based Text Steganography Method, in: Journal of Systems and Software, Vol. 85, Issue. 10, pp. 2385–2394.