پنهان نگاری به شیوه LSB بر پایه ی آنالیز داده

استاد: ساجد محیسن[۱] ، دانشجو: نعیم عبادی*[۲]، دبیر: زهرا تلیان[۳]

۱- استاد موسسه آموزش عالی خرد
sajednet@gmail.com
۲- دانشجوی موسسه آموزش عالی خرد
nmebadi1361@gmail.com
۳- معلم آموزش و پرورش
toliyanteacher92@gmail.com

چکیده

پنهان نگاری شیوه ای امن و مطمئن برای ارسال اطلاعات به شمار می آید که در آن پیام ارسالی درون یک رسانه پوششی مخفی شده و در قالب محموله ای واحد به مقصد ارسال میگردد. از آنجا که داده های ارسالی و رسانه پوششی باید سازگاری نسبی داشته باشند همچنان روش هایی برای ایجاد انعطاف بیشتر در سازگاری این دو مطرح میگردد. هدف از این مقاله ایجاد روشی برای ذخیره بیشترین حجم ممکن از اطلاعات به همراه امنیت ادغام شده در خود روش ذخیره سازی به طور غیر قابل تفکیک و توانایی مقاومت مطلوب در برابر حملات پنهان شکنی است. تاکنون ایده های زیادی ارائه شده است که هر کدام مزایا و معایب خود را دارد. در اینجا سعی شده تا جایی که ممکن است کاربر بتواند آزادانه رسانه پوششی را از لحاظ حجمی و ابعاد تصویر انتخاب کند و بکار ببندد برای این منظور باید در قالب ذخیره سازی اطلاعات مخفی تغییراتی را اعمال کرد و همچنین از لحاظ امنیت جوابگوی انتظارات کاربر باشد. برای رسیدن به هدف از شیوه پنهان نگاری LSB همراه با آنالیز داده استفاده شده که جنبه های بهینه سازی حجم اطلاعات و امنیت آن را فراهم می کند . همچنین کمترین تغییر در رسانه پوششی را به دنبال خواهد داشت.

کلمات کلیدی: استگانوگرافی ، LSB ، بهینه سازی ، امنیت ، داده

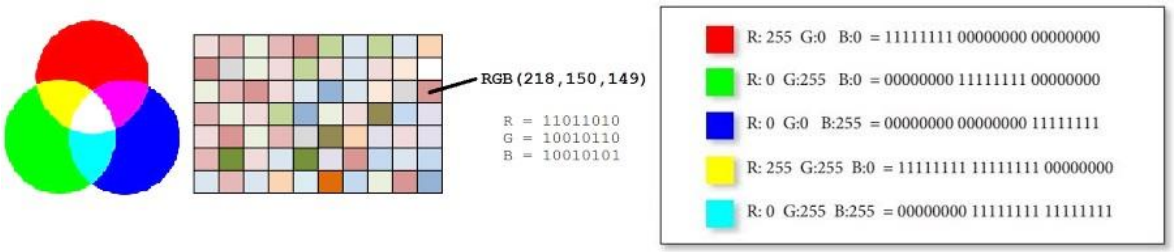
۱–مقدمه

تبادل اطلاعات در جوامع امروزی امری ضروری و اجتناب ناپذیر است. از آنجا که پیشرفت و ترقی جوامع مرهون کسب اطلاعات بیشتر و دانش حاصل از آن است، هر جامعه ای سعی در افزایش حجم اطلاعات خود و پیشی گرفتن از جوامع دیگر دارد. این امر در سطح فردی نیز صدق میکند. پیشتاز بودن در عرصه های دانش منجر به بروز دانش انحصاری و قلمرو اطلاعاتی افراد و سازمان ها شده که باید از رخنه کردن بیگانه در آن جلوگیری کرد. و این خود بحث امنیت اطلاعات را پیش می کشد که با وجود ارتباطات و نیاز به رد و بدل کردن و جابجایی اطلاعات باید آنها را از دسترس دیگران محفوظ بداریم[1]. استگانوگرافی(پنهان نگاری) می تواند در این زمینه کارساز باشد و راه کارهایی را برای مخفی کردن پیام ها و اطلاعات ارسالی به ما ارائه دهد در واقع استگانوگرافی هنری است که می کوشد پیامی را در پوششی رسانه ای و عمومی جاسازی کند به گونه ای که در حالت معمولی اثری از پیام دیده نشود و شبهه ای را ایجاد نکند. در این نوع از پنهان نگاری پیام پس از پنهان شدن در رسانه پوششی توسط فرستنده ارسال و به گیرنده رسیده سپس توسط متدی که در بین دو طرف

مقرر شده از حالت پنهان خارج شده و آشکار می شود[۲] . این مقاله سعی دارد با زبانی ساده روش بهینه سازی پنهان نگاری بر پایه آنالیز داده را توضیح دهد.

۲– فشرده سازی و بهینه سازی ذخیره اطلاعات

در راستای حفظ اطلاعات محرمانه هر اقدام هرچند ناچیز امکان موفقیت و رسیدن به هدف را افزایش می دهد . هر مقدار امکانات بیشتر برای استفاده مهیا کنیم و انعطاف بیشتری به روش کار خود دهیم محدوده کاربردی خود را افزایش داده و اقبال بیشتری خواهیم داشت. از این رو در این مقاله سعی شده است با استفاده از روش ذخیره سازی فشرده و رمزگذاری داده ها به شیوه LSB و ادغام آن با آنالیز متنی و تفکیک کاراکتری بر مبنای شخصی سازی مجموعه کاراکترهای استفاده شده در پیام ارسالی تا آنجا که امکان دارد در کاهش حجم ذخیره سازی و ارتقای امنیت داده ها تغییراتی ایجاد کنیم [۳] . از آنجا که هر پیکسل از نوع RGB شامل سه طیف رنگی است میتوان از کدهای هر طیف که شامل هشت بیت است یک بیت که دارای کمترین ارزش است را بکار گیریم این مناسب ترین روش استگانوگرافی در شیوه ی LSB می باشد که کمترین تغییر را در رسانه پوششی ایجاد میکند که قابل چشم پوشی می باشد[۴]. (شکل شماره ۱)



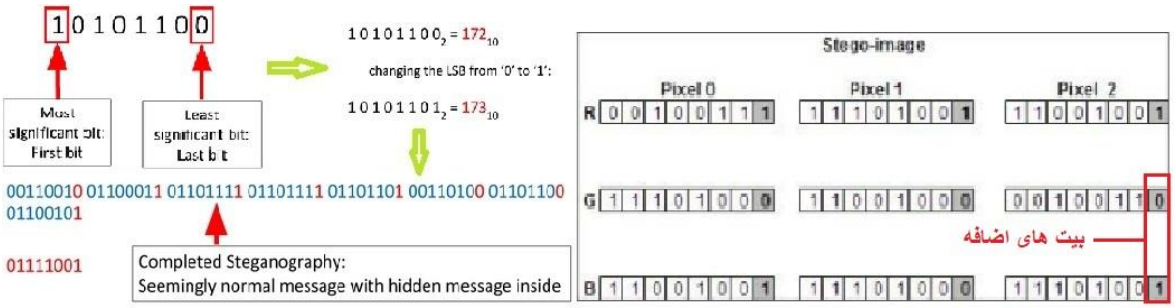
شکل شماره (۱) طیف های رنگ RGB در تصاویر

بنابراین هر پیکسل توانایی نگهداری سه بیت را برای ما فراهم میسازد. برای ایجاد قالب فشرده شده از اطلاعاتی که نیاز به پنهان سازی دارند ابتدا باید ببینیم تنوع کاراکترهای استفاده شده در پیام ارسالی چه مقدار است. همینطور که میدانید مجموعه پرکاربرد حروف زبان انگلیسی، ارقام و کاراکترهای عمومی دیگر ۱۲۸ کاراکتر است که با اعداد ۰ تا ۱۲۷ میتوان آنها را کدبندی نمود. بنابراین حداکثر بیت های لازم جهت کدبندی در مبنای دودویی هفت بیت می باشد[۴]. به این ترتیب برای کدبندی هفت بیتی در بیت های کم ارزش طیف های نوری پیکسلهای رسانه پوششی به ازای هر کاراکتر نیاز به حداقل سه پیکسل می باشد پس حجم ذخیره داده ها طبیعی و بدون فشرده سازی خواهد بود. در راستای کاهش حجم و افزایش ظرفیت ذخیره سازی ابتدا تا جایی که ممکن است تعداد بیت های کدبندی داده ها را کاهش می دهیم[۴]. برای این منظور تنوع کاراکترهای استفاده شده در پیام انتقالی را شمرده و کاراکترهای استفاده شده در پیام ارسالی را در مجموعه ای جدا به هدف ایجاد منبع خلاصه شده از کاراکترها جهت کاهش کدهای تشخیص کاراکتر مورد نیاز با اشاره به جایگاه آن در مجموعه بصورت خلاصه شده، نگهداری می کنیم و بستگی به طول مجموعه کاراکتری منبع ایجاد شده برای کدبندی کاراکترهای آن پیام از دسته های بیتی هفت ، شش ، پنج و چهار استفاده می کنیم. برای درک بهتر این موضوع باید به آمار استفاده از کاراکترهای حروف و اعداد و علامت های موجود در زبان لاتین انگلیسی مراجعه کرد بدیهی است در اکثر پیام های رد و بدل شده درصد زیادی از کاراکترها بکار برده نمی شوند حروف بزرگ انگلیسی درصد استفاده کمتری دارند و از برخی علائم به ندرت استفاده می شود. کدبندی هفت بیتی می تواند محدوده کاراکترهای زبان لاتین را پوشش دهد که تعداد ۳۲ کد ابتدایی اسکی غیر قابل چاپ و بدون کاربرد است. با توجه به ندرت استفاده از برخی کاراکترهای دیگر در صورت کاهش تعداد تنوع کاراکترها تا مرز ۶۴ کاراکتر میتوانیم برای بهینه سازی حجم ذخیره سازی اطلاعات از کدبندی

شش بیتی استفاده کنیم. همچنین در صورت کاهش مقدار تنوع کاراکترهای پیام تا ۳۲ کاراکتر متنوع، استفاده از کدبندی پنج بیتی توصیه می شود این بهینه سازی تا مرز چهار بیتی ادامه دارد. حتی کدبندی با دسته های بیتی چهارتایی می تواند ۱۶ حالت مختلف از تنوع کاراکترها را پوشش دهد و این امر زمانی که پیام ارسالی از کلیدهای عددی که به هدف رمزگشایی بسته های خاص مورد استفاده قرار میگیرد از کاراکترهایی همچون ده کاراکتر ارقام و پنج کاراکتر دیگر مانند ممیز، نقطه، فاصله و ... تشکیل شده باشد مناسب است و باعث کاهش فضای هدر رفته در روش پنهان سازی LSB می شود.

۳–نگاشت بیتی کاراکترها

نگاشت بیتی نوعی بهینه سازی است که به ما اجازه می دهد وجود یا عدم وجود یک کاراکتری که شناخته شده می باشد توسط یک بیت ارزیابی شود که از بیت های اضافه برای این منظور استفاده می کنیم. منظور از بیت های اضافه بیت هایی هستند که بعد از ثبت بسته کدبندی مثلا هفت بیتی که بیت هفتم آن در بیت اول پیکسل شماره سه ذخیره می شود و بیت های شماره دو و سه خالی و بدون استفاده باقی می مانند می توانیم از بیت شماره دو از پیکسل سوم در زمانی که کاراکتر بعدی تکرار کاراکتر فعلی است مقدار یک را وارد کنیم در این صورت هنگامی که کار رمزگشایی انجام میشود با خواندن بیت شماره هشت(بیت دوم از پیکسل سوم) کاراکتر فعلی را مجددا برای چاپ به خروجی میفرستد و از تکرار کدبندی جلوگیری می شود این امر در زبان انگلیسی شایع است و خود می تواند مقدار قابل توجهی از فضای ذخیره سازی را بهینه کند. بعنوان مثال تکرار ارقام مشابه در کنار هم امری عادی و همچنین تکرار حروف در کلماتی همچون Green، Yellow، www.google.com، 336557 و موارد مشابه دیگر رایج است. همینطور بیت شماره نه(بیت سوم از پیکسل سوم) در صورتی که کاراکتر بعدی فاصله بین کلمات می باشد مقدار یک را دریافت می کند و در نگاشت فاصله ها تقریبا هیچ فضای اضافی مورد نیاز نیست. برای جلوگیری از تکرار ثبت اضافی فاصله بین کلمات در صورتی که بصورت مکرر و پشت سر هم در پیام ارسالی وجود داشته باشد(در حالتی که کاراکتر بعدی هم تکرار کاراکتر فعلی و هم فاصله باشد) از یک دستور شرطی استفاده می کنیم به صورتی که بیت شماره هشت مقدار یک را دریافت کرده باشد از تغییر در بیت نهم جلوگیری شود. با سنجش درصد فاصله ها نسبت به متن های عمومی میتوانیم این اقدام را یکی از تاثیرگذارترین راه های ممکن برای بهینه سازی شمرد در بسته کدبندی هفت بیتی هر فاصله بین کلمات پیام موجب صرفه جویی در مصرف بیت های کم ارزش سه پیکسل می شود و تعداد نه بیت را برای کاراکترهای بعدی تامین می کند.(شکل شماره ۲)



شکل شماره (۲) کدبندی داده ها در بیت های کم ارزش

تعداد بیت های بسته کدبندی در استفاده از نگاشت بیتی اهمیت زیادی دارد این نوع بهینه سازی در کدبندی هفت، پنج و چهار بیتی قابل استفاده است که به ترتیب تعداد بیت های اضافه در آن ها دو، یک و دو می باشد. کدبندی پنج بیتی می تواند از یک بیت اضافه بهره ببرد که به نظر می آید استفاده از آن برای هر کدام از ثبت فاصله یا تکرار کاراکترها وابسته به نوع متن پیام است که در نهایت این دو با هم رقابتی نزدیک دارند و میتوان یکی را انتخاب کرد. در کدبندی های کوتاه تر از

هفت بیت تعداد پیکسل هایی که برای ذخیره هر بسته کد داده استفاده می شود دو پیکسل است بدین ترتیب فضای ذخیره سازی مقرون به صرفه ای را فراهم می کند که با ادغام آن با مزایای نگاشت بیتی کارایی بالاتری خواهد داشت.

۴– تامین امنیت همراه با کدبندی بیتی کوتاهتر

در هنگام استفاده از بسته های کدبندی کوچک سازی شده نیاز است که محدوده کاراکتری داده های پیام بطور کامل پوشش داده شود و لازمه این امر شخصی سازی کدهای کاراکترهای موجود در پیام ارسالی و اعتبار دادن به صورت فشرده و خلاصه سازی منبع حروف و علامت ها می باشد به عنوان مثال در این مرحله برای اینکه بتوانیم کاراکتری مانند Z با کد اسکی ۹۰ را در کدبندی کوتاه تر از هفت بیتی نمایش دهیم پس از اطلاع از مقدار تنوع کاراکتری موجود در پیام ارسالی، با درج تعداد و سپس مجموعه کاراکترهای موجود در پیام به ترتیب دریافت و بدون تکرار در ابتدای رسانه پوششی با کدبندی بسته ای هشت بیتی و با استفاده از روش LSB می توان خود مجموعه کاراکتری پیام را به همراه رسانه پوششی انتقال داد و سپس برای پنهان نگاری پیام ارسالی و کاهش حجم ذخیره سازی بجای کدبندی خود کاراکترها از کدبندی موقعیت کاراکترها در مجموعه استفاده می کنیم بنابراین برای نمایش حرف Z با کد اسکی ۹۰ کافیست موقعیت آن در مجموعه را کدبندی کنیم به فرض اگر تنوع مجموعه کاراکترهای یک پیام ۳۲ باشد تنها با کدبندی بسته ای پنج بیتی امکان اشاره به کاراکترها و رسیدن به هدف فراهم می شود این کار باعث کاهش طول بسته های کدبندی شده می گردد و کاهش حجم را به دنبال دارد در نهایت با روند رمزگشایی و رسیدن به کد مربوط به محل قرارگیری هر کاراکتر و مراجعه به آن در ابتدای پیام مخفی به روش LSB می توان آن را بازیابی کرد و برای نمایش پیام مخفی شده استفاده نمود. (شکل شماره ۳)

شکل شماره (۳) روش ثبت چهار بیتی در مجموعه های کوتاه

طول مجموعه: محل ذخیره تعداد کاراکترهای مجموعه منبع با کدبندی هشت بیتی(به شیوه صعودی که در ادامه به آن خواهیم پرداخت)

کاراکتر با رنگ سیاه: کاراکترهای مجموعه منبع با کدبندی هشت بیتی

کدهای با رنگ سفید: اندیس های کاراکترهای مجموعه که محل قرار گیری و ترتیب کاراکترها را نشان می دهد

کدهای رنگ قرمز: همان داده باینری که به کاراکترهای مجموعه اشاره می کند

خانه زرد رنگ (شکل ۳) نشان می دهد که می توان با کدبندی عدد پنج که اندیس کاراکتر Q(با کد اسکی ۸۱) در بسته چهار بیتی به کاراکتری با کد اسکی بیشتر از شازده اشاره کرد که موجب کاهش حجم در ذخیره سازی می گردد. استفاده از این روش نه تنها باعث کاهش حجم کدبندی و بسته های بیتی میگردد بلکه یک لایه ی امنیتی نیز به اطلاعات مخفی شده اضافه می کند که باعث مقاومت در برابر پنهان شکنی و دشواری استخراج اطلاعات استگانو شده در تصویر می گردد با وجود سفارشی کردن کاراکترها در صورت خوانده شدن بیت های کم ارزش رسانه پوششی به مجموعه ای از کاراکترهای نامفهوم و نا مرتبط برخورد کرده و بدون رمزگشایی بیتی مخصوص و ترکیب آن با مجموعه منبع برای بدست آوردن موقعیت مکانی

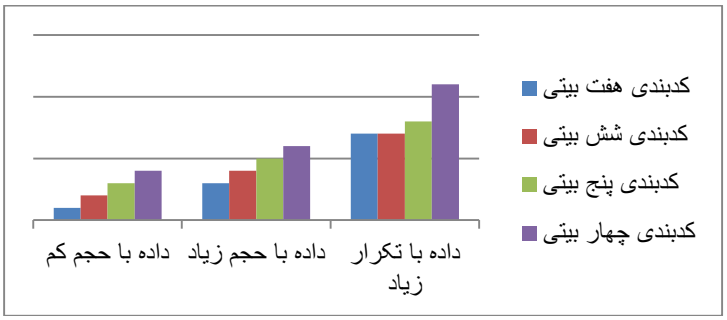
هر کاراکتر ، بازیابی پیام مخفی شده غیر ممکن خواهد بود همچنین درج کاراکترهای پیام ارسالی بدون ترتیب صورت گرفته و این خود به بالا بردن سطح امنیت در پنهان نگاری کمک میکند.

۵– کد بندی صعودی پیکسل ها

علاوه بر روش مذکور برای رسیدن به سطح امنیتی بالاتر میتوان ترتیب کدبندی پیام ارسالی را با مرتب سازی صعودی پیکسل های رسانه پوششی با احتساب هفت بیت پرارزش و پس از به صفر رساندن بیت های کم ارزش انجام داد بدین صورت داده ها از ابتدای رسانه پوششی ذخیره نمی شوند و در تمامی نقاط تصویر ذخیره سازی میشوند و چون بدون احتساب بیت های کم ارزش مرتب سازی شده اند پس از مقدار دهی بیت های کم ارزش ترتیب آنها حفظ شده و دچار تغییر نمی شوند که هنگام رمزگشایی نیز با احتساب هفت بیت پر ارزش میتوان به ترتیب اولیه پیکسل ها و شیوه ذخیره داده ها دسترسی داشت در این روش پیکسل های با مقادیر یکسان را بر حسب اولویت مکانی رتبه بندی میکنیم و با تبعیت از این شیوه در رمزگشایی و بازیابی داده ها مشکلی نخواهیم داشت. (شکل شماره ۳)

۶– نتیجه گیری

در شیوه ارائه شده توانسته ایم حجم داده های ارسالی را بدون هیچ کم و کاستی در اطلاعات به مقدار زیادی کاهش دهیم. مقدار صرفه جویی در فضای ذخیره سازی تایع تنوع کاراکتری در پیام ارسالی می باشد در این روش هیچ گونه محدودیتی برای نگاشت متنی وجود ندارد بلکه این خود شیوه ی پنهان نگاری است که با توجه به تنوع در استفاده از کاراکترها در پیام تصمیم میگیرد که از کدام طول بسته بیتی مناسبتر خواهد بود. با استفاده از این روش هر مقدار که تنوع کاراکترها کمتر باشد و میزان اطلاعات ارسالی بیشتر ، درصد تاثیر در بهبود ذخیره سازی عملکرد بهتری را از خود نشان می دهد. همچنین با توجه به تعداد اندک کاراکترهایی که بعنوان مجموعه منبع به همراه داده ها در رسانه پوششی ارسال می شود اختلاف چندانی در حجم داده ها ایجاد نمی کند و میتوان آن را نادیده گرفت. نگاشت تک بیتی کاراکترهای تکراری و فاصله های بین کلمات کارایی بالایی در کاستن از حجم داده و بالا بردن سرعت کدبندی اطلاعات و همچنین رمزگشایی نهایی پیام ارسالی از خود نشان می دهد.(شکل شماره ۴)



شکل شماره (٤) چارت آماری رابطه طول بسته های بیتی و حجم داده با تکرار داده ها

وجود مجموعه کاراکتری منبع به همراه پیام ارسالی باعث شخصی سازی و به گونه ای رمزنگاری کاراکترهای داده میگردد که دلیل کوتاه بودن بسته های بیتی را توجیه می کند. امنیت اطلاعات در این شیوه با فشرده سازی کدبندی داده ها تامین می شود که خود لایه ای محافظ در برابر حملات پنهان شکنی است. علاوه بر این روش انتخاب پیکسل ها برای ثبت داده الگویی صعودی در رسانه پوششی ایجاد می کند که داده ها را در نقاط مختلف تصویر توزیع کرده و از انباشتگی چگالی در

ابتدای رسانه جلوگیری میکند. بدیهی است که نیاز به پردازش رسانه پوششی در سرعت عملکرد پنهان نگاری تاثیر منفی را در پی دارد این نکته می تواند یکی از معایب روش بیان شده باشد. در مجموع مزایای فشرده سازی و امنیت ارائه شده غیر قابل انکار و چشم پوشی است. کاربرد این شیوه ارسال اطلاعات بیشتر با استفاده از رسانه های پوششی با حجم کمتر و سطح امنیتی مطلوب را برای ما فراهم می کند.

۷–مراجع

۱. مرادی, هادی و محسن سرداری زارچی، ۱۳۹۷، ارایه مکانیزمی مبتنی بر نگاشت آشوب لجستیک و استگانوگرافی در تصاویر دیجیتال، کنفرانس ملی تحقیقات نوین در مهندسی کامپیوتر برق فناوری اطلاعات، مبارکه، دانشگاه آزاد اسلامی واحد مبارکه،
https://www.civilica.com/Paper-ECIT01-ECIT01_003.html

۲. موسوی, سید محمد مهدی و یوسف ترابی گل سفید، ۱۳۹۷، مروری بر روش های نهان نگاری در تصاویر، کنفرانس بین المللی تحقیقات بین رشته ای در مهندسی برق، کامپیوتر، مکانیک و مکاترونیک در ایران و جهان اسلام، کرج، دانشگاه جامع علمی کاربردی سازمان همیاری شهرداری ها،
https://www.civilica.com/Paper-ECMM01-ECMM01_065.html

۳. جلالی باروق, مریم و پیمان ایوبی، ۱۳۹۶، استگانوگرافی ویدیوی دیجیتال براساس دنباله های فرکتالی آشوبناک، دومین کنفرانس ملی ریاضی:مهندسی پیشرفته با تکنیک های ریاضی، ارومیه، دانشگاه آزاد اسلامی واحد ارومیه،
https://www.civilica.com/Paper-MAEMT02-MAEMT02_118.html

٤. افسرده, حمید و مرضیه دادور، ۱۳۹۴، استگانوگرافی به همراه کریپتوگرافی و فشرده سازی برپایه شبکه های عصبیMLP ، کنفرانس بین المللی یافته های نوین پژوهشی درمهندسی برق و علوم کامپیوتر، تهران، موسسه آموزش عالی نیکان،
https://www.civilica.com/Paper-COMCONF01-COMCONF01_548.html