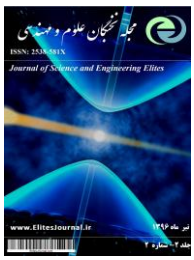
www.ElitesJournal.ir

مجله نخبگان علوم و مهندسی

Journal of Science and Engineering Elites

ISSN 2538-581X

جلد ۲- شماره ۲- سال ۱۳۹۶



مخفی کردن اطلاعات در تصاویر متنی با استفاده از پیکسل‌های موجود روی لبه‌های حروف فارسی

وحید یزدانی[۱]، موسی فلاح خورسند[۲*]

۱- کارشناسی ارشد فناوری اطلاعات از دانشگاه شاهد تهران

۲- کارشناسی ارشد مهندسی برق ، باشگاه پژوهشگران جوان و نخبگان دانشگاه آزاد اسلامی واحد نور

*afshin.khorsand@gmail.com

ارسال: فروردین ماه ۹۶ پذیرش:تیر ماه ۹۶

چکیده

یکی از مسائل مهم در مبادله پیام‌های اینترنتی، مسئله امنیت تبادل اطلاعات است. واترمارکینگ یکی از مهمترین روش‌های مخفی-سازی اطلاعات است که برای محافظت از حق مالکیت از محتوی چند رسانه‌ها استفاده می‌شود. در این میان در زمینه واترمارکینگ متون فارسی یا عربی با استفاده از خصوصیات حروف به کاررفته در متن به نسبت متون انگلیسی و چینی تحقیقات کمتری انجام شده است. هدف از روش ارائه شده در این مقاله که برای واترمارکینگ متون فارسی پیشنهاد می‌شود، عدم مشاهده‌پذیری و در نتیجه امنیت بالای آن می‌باشد. روش پیشنهادی از زوایای پیکسل‌های موجود در لبه حروف "ک" و "گ" برای مخفی‌سازی اطلاعات استفاده می‌کند. این زوایا که پس پردازش اولیه حروف و اعمال عملگرهای لبه‌یابی روی حروف بدست می‌آیند، منحصر بفرد بوده و از آن برای انتخاب نقطه مورد نظر و کلیدی جهت تعبیه اطلاعات استفاده می‌شود. روش پیشنهادی که جزء زیر مجموعه روش های کور می باشد، از عدم مشاهده پذیری بالاتری نسبت به روش های قبلی برخوردار بوده و گیرنده قانونی متن واترمارک شده تنها با در اختیار داشتن الگوریتم واترمارکینگ مناسب قادر به استخراج خواهد بود و این باعث امنیت بالای متن واترمارک شده طبق روش پیشنهادی می‌شود.

کلمات کلیدی: واترمارکینگ، مخفی‌سازی اطلاعات، متون فارسی، خط پایه.

۱. مقدمه

استفاده از کامپیوتر، شبکه‌های کامپیوتری، اینترنت و نقل وانتقال انواع اطلاعات و رسانه[۱] دیجیتال اعم از صوت، تصویر، ویدئو و نرم افزار از طریق این شبکه‌های اطلاعاتی، امری غیر قابل اجتناب در دنیای دیجیتال امروزیست. بنابراین تصدیق هویت رسانه، امنیت اطلاعات در برابر جعل وکپی برداری غیرمجاز، انتقال محرمانه اطلاعات و ... همگی از مواردی به شمار می‌روند که در دنیای دیجیتال امروزی از اهمیت بالایی برخوردارند[۱].

[1] Media

در این میان تکنیک‌های پنهان نگاری[1] و واترمارکینگ به عنوان مهمترین روش‌های پنهان‌سازی اطلاعات محسوب می‌شوند. واترمارکینگ می‌تواند به عنوان یکی از روش‌های پنهان‌نگاری دیده شود که تمرکز بیشتر بر روی استحکام و کمتری بر روی امنیت دارد. امنیتی که پنهان نگاری و واترمارکینگ تامین می‌کنند، با امنیت تامین شده توسط رمزنگاری[2] متفاوت است. امنیت واترمارکینگ توسط عدم آگاهی دیگران از وجود اطلاعات مخفی شده در رسانه، تامین می‌شود ولی در رمزنگاری اطلاع از این که در رسانه اطلاعاتی مخفی است وجود داشته، و کلیه افرادی که دارای کلید باشند قادر به رمزگشایی می‌باشند. بر خلاف واترمارکینگ در رمزنگاری هیچ محافظتی از محتوی پیام پس از رمزگشایی ندارد. نکته دیگر تفاوت در فرآین شکستن سیستم‌های رمزنگاری و واترمارکینگ است. شکستن سیستم‌های رمزنگاری زمانی اتفاق می‌افتد که حمله کننده بتواند پیام مخفی را بخواند ولی در واترمارکینگ این اتفاق زمانی روی می‌دهد که:

۱-حمله کننده حضور واترمارکینگ را تشخیص دهد

۲-حمله کننده قادر به خواندن، اصلاح یا حذف پیام مخفی باشد.

در حقیقت واترمارکینگ یک روش برای محافظت از حق مالکیت[3] از محتوی چند رسانه‌ها می‌باشد. این چند رسانه‌ها، رسانه-های مختلفی از جمله متن، تصویر، ویدئو، صدا و اشیاء گرافیکی را نمایش داده که هرکدام ویژگی‌های مخصوصی دارند ازاین ویژگی‌ها برای مخفی‌سازی اطلاعات، دردرونشان استفاده می‌شود. از این رو لازم است که الگوریتم‌های واترمارکینگ مخلتفی متناسب با نوع اطلاعات توسعه داده شود[۲] [۳].

در میان رسانه‌ها، اسناد متنی خواص مختص به خود را دارند. طبیعت باینری، ساختار بلوک/خط/کلمه، تفکیک مشخص بین پیش‌زمینه و پس‌زمینه که از جمله آن‌هاست. الگوریتم‌های نشان گذاری نیز از این ویژگی‌ها برای مخفی‌سازی اطلاعات استفاده می‌کنند[۴]. برخلاف واترمارکینگ متنی، روش‌های بسیار زیاد و متنوعی برای نشان گذاری دیگر انواع رسانه (صوت، تصویر(رنگی و خاکستری) و ویدئو) وجود دارد. این در حالی است که در زمینه واترمارکینگ متن فارسی با استفاده از خصوصیات ویژه آن، همانند متون چینی و انگلیسی تحقیقات چندانی صورت نگرفته است[۱].

به همین دلیل در این مقاله روش جدیدی برای واترمارکینگ متون فارسی پیشنهاد می‌شود که از زوایای پیکسل‌های موجود در لبه حروف "ک" و "گ" برای مخفی‌سازی اطلاعات استفاده می‌کند. لازم به ذکر است که هدف از انتخاب دو حرف "ک" و "گ" استفاده از ویژگی منحصر بفردی در ماتریس زوایه است که پس از پردازش اولیه این حروف مشخص شده و استفاده از نقطه منحصر بفرد که در مسیر سرکش به بدنه اصلی این حروف وجود دارد، می‌باشد. استفاده از روش پیشنهادی در مابقی حروف فارسی بشرط پیدا کردن ویژگی منحصر بفرد در ماتریس زاویه ممکن بوده و به هیچ وجه نام حروف ملاک اصلی انتخاب در جهت استفاده از روش پیشنهادی نیست. مابقی مقاله به صورت زیر سازماندهی شده است:

دربخش دوم، الگوریتم‌های واترمارکینگ دسته‌بندی شده و پارامترهای مؤثر در ارزیابی این الگوریتم‌های بحث خواهد شد. بخش سوم مربوط به کارهای انجام شده در زمینه واترمارکینگ فارسی و عربی بوده، در بخش چهارم الگوریتم واترمارکینگ پیشنهادی شرح داده خواهد شد. بخش آخر شامل ارزیابی و نتیجه‌گیری روش پیشنهادی خواهد بود.

۲. دسته‌بندی و پارامترهای مؤثر در ارزیابی الگوریتم‌های واترمارکینگ

در ارزیابی الگوریتم‌هایی که برای ایجاد و کشف واترمارکینگ استفاده می‌شوند، پارامترهای زیر متداول است:

[1] Steganography
[2] Cryptography
[3] Copyright

۱- استحکام[1]: مربوط به نحوه جا ساری اطلاعات به طوری که در برابر حملات پایدار باشد. این پارامتر هرچه بالاتر باشد نشانه آن است که میزان تخریب و اثرپذیری واترمارک در اثر تغییرات دررسانه میزبان کمتر است.

۲- ظرفیت[2]: این پارامتر مشخص می کند که حداکثر حجم ممکن که رسانه برای واتر مارک قادر است فراهم کند چقدر است.

۳- امنیت: تعیین می‌کند که رسانه میزبان تا چه میزان وجود واترمارک را می‌تواند مخفی نگه داشته و باعث جلوگیری از دستکاری، جعل و حذف پیام شود[۵].

روش‌های واترمارک را به صورت‌های زیر می‌توان دسته‌بندی کرد:

۱- بر اساس قابل مشاهده بودن بیت‌های واترمارک:

الف) قابل مشاهده[3]: واترمارک‌های تعبیه شده در هنگام رویت محتوی رسانه با یک روشی قابل دیدن هستند.

ب) مرئی[4]: با چشم قابل دیدن نبوده ولی با یک الگوریتم رمزگشایی مناسب می‌توان دید.

۲- بر اساس محتوی یا نوع رسانه میزبان:

الف) واترمارکینگ تصاویر

ب) واترمارکینگ ویدئو

پ) واترمارکینگ صدا

ج) واتر مارکینگ متن

۳- بر اساس نیاز به سند اصلی هنگام استخراج :

الف)کور[5] : در هنگام فرایند استخراج نیاز به سند اصلی نیست.

ب) غیرکور[6] : در هنگام فرایند استخراج نیاز به سند اصلی است.

۴- بر اساس میزان استحکام:

الف) واترمارکینگ مقاوم[7]: تغییر یا اصلاح محتوی رسانه تاثیری روی بیت‌های واترمارک ندارد.

ب) واترمارکینگ ضعیف[8]: تغییر یا اصلاح محتوی رسانه بر روی بیت‌های واترمارک اثر می گذارد[۳].

۳. مروری بر مطالعات پیشین

متون زبان فارسی بر خلاف متون انگلیسی از راست به چپ نوشته می‌شوند. زبان فارسی دارای ۳۲ حرف بوده که ۱۸ حرف از ۳۲ حرف آن دارای ۱ تا ۳ نقطه می‌باشند. نقاط ممکن است در بالای حرف مانند «ت» و یا در پایین حرف مانند «پ» و یا در میان حرف مانند «چ» قرار بگیرند در حالیکه در زبان انگلیسی فقط دو حرف «i» و «j» دارای نقطه می‌باشند. بر خلاف زبان انگلیسی حروف در کلمات فارسی به هم پیوسته بوده و با توجه به مکان قرارگیری حرف در کلمه ممکن است شکل‌های متفاوتی داشته باشد. به عنوان مثال حرف "ع" بسته به مکان قرارگیری در کلمه به صورت‌های "ع"، "ـعـ"، "عـ" و "ع" نوشته می‌شود[۶].

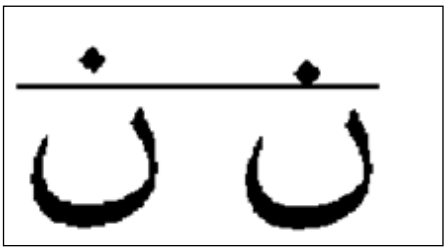
با توجه به خصوصیات ذکر شده درباره حروف فارسی کارهای کمی در زمینه واترمارکینگ فارسی انجام شده است. از جمله کارهای انجام شده در زمینه واترمارکینگ فارسی و عربی می‌توان به موارد زیر اشاره کرد:

---

[1] Robustness
[2] Data rate
[3] Visible
[4] Invisible
[5] Blind
[6] Informed
[7] Robust
[8] Fragile

۱.۳. روش نقطه

برای درج بیت صفر مکان نقطه تغییری نکرده و برای درج بیت یک نقطه کمی به سمت بالا انتقال می یابد. شکل ۱ نمونه‌ای از این روش تعبیه اطلاعات را نمایش می‌دهد[۷].



شکل۱- نمونه ای از واترمارکینگ با جابجایی نقاط [۷]

۲.۳. روش استفاده از حروف با کدهای متفاوت

بعضی حروف مثل "ک" و "ی" در عربی و فارسی به دو شکل متفاوت نوشته می شود. برای درج بیت صفر از شکل فارسی حرف "ی" و "ک" استفاده شده و برای درج بیت یک از شکل فارسی حرف "ی" و "ک" استفاده می‌شود[۸].

۳.۳. روش کاراکتر کشیده

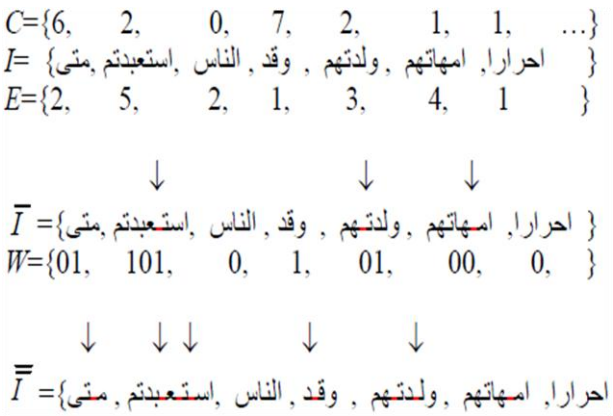
الف) برای درج بیت صفر کاراکتر کشیده به قبل یا بعد از حرف فاقد نقطه اضافه شده و برای درج بیت یک کاراکتر کشیده به قبل یا بعد از حرف دارای نقطه اضافه می‌شود[۲].

ب) در روشی دیگر ابتدا از یک کلید محرمانه به عنوان هسته اعداد تصادفی به صورت رابطه زیر ایجاد می‌شود:

$$ci \in [0,7] \text{و} C=\{c0, c1, c2 \ldots cn\}, \quad (1)$$

که n برابر با تعداد کلمه در متن است

مجموعه ای به نام I که بیانگر مجموعه کلمات موجود در متن ومجموعه ای به نام E که معرف تعداد کارکتر قابل کشیده در هر کلمه است، تعریف می‌شود. درهر کلمه اگر مقدار C کمتر از مقدار E باشد، یک کارکترکشیده درجایی که C اشاره می کند درج شده و مجموعه حاصله را Ì نامیده می‌شود (برای گمراهی).حال مقادیری که قرار است مخفی شود در Ì درج شده و Ì نامیده می‌شود. در شکل ۲ این روش روی برخی از کلمات اعمال شده است[۹].



شکل۲- روش استفاده از کاراکتر کشیده با استفاده از کلید[۹]

۴.۳. روش یونیکد

در این روش از یک کلمه یکسان که دارای دو شکل ظاهری متفاوت بوده و در سیستم استاندارد یونیکد، دو کد متفاوت دارد، برای درج بیت‌های واترمارک استفاده می‌شود. به عنوان مثال در کلمه لا برای درج بیت یک از " لـا " و برای درج بیت صفر از شکل عادی، یعنی "لا" استفاده خواهد شد.

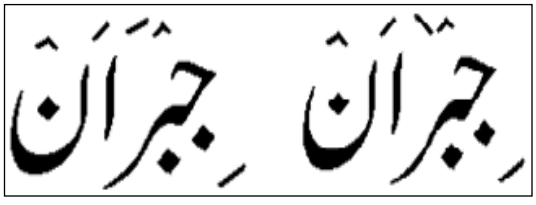
کلمه "لا" دارای کد FEFB و کلمه " لـا" دارای کد 0640 در سیستم استاندارد یونیکد است[۱۰].

۵.۳. روش شبه فاصله

از آنجایی که در زبان فارسی از نیم فاصله در برخی از کلمات استفاده می شود (مانند فاصله بین "می" و "باشد" در کلمه "می‌باشد") می‌توان از آن برای درج بیت‌های واترمارکینگ استفاده کرد به طوری که در هنگام برخورد به نیم فاصله به صورت زیر عمل می‌شود: برای درج بیت صفر به نیم فاصله چیزی اضافه نکرده و برای درج بیت یک به نیم فاصله یک فاصله اضافه خواهد شد[۶].

۶.۳. روش استفاده از علائم موجود بر حروف

الف) برای تلفظ صحیح کلمات در زبان‌های عربی و اردو از اعراب(فتحه، کسره، ضمه) استفاده می شود که در این بین فراوانی تعداد فتحه بیشتر از دو تای دیگر است و می‌توان از آن برای برای درج بیت‌های واترمارکینگ استفاده کرد به طوری که برای درج بیت صفر از شکل عادی فتحه استفاده شده و برای درج بیت یک از شکل معکوس شده فتحه استفاده می‌شود. در شکل زیر فتحه موجود بر "ر" در کلمه سمت راست معکوس شده است. شکل ۳ اعمال این روش روی کلمات دارای علامت را نشان می‌دهد[۱۱].



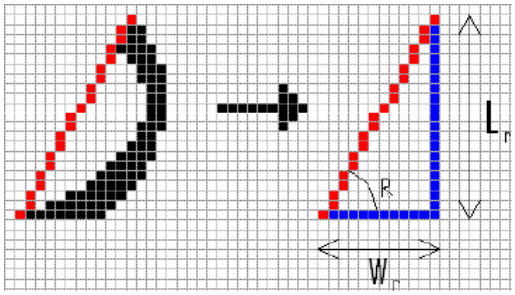
شکل۳- روش استفاده از فتحه معکوس [۱۱]

ب) در روش دیگر از فتحه برای درج بیت یک و کسره و ضمه برای درج بیت صفر استفاده می‌شود. هرچند ممکن است به علت درج بیت‌های یک و یا صفر متوالی و عدم تناسب علائم موجود بر حروف با مقدار بیت ذخیره شده، از همه علائم استفاده بهینه نشود[۱۲].

پ) در روش سوم سعی شده است که مشکل ذکر شده در روش فوق حل شود. بدین صورت که برای درج بیت یک علائم موجود برحرف (فتحه، کسره، ضمه و ساکن) در سرجای خود باقی مانده و برای درج بیت صفر این علائم از حروف برداشته می‌شود. هرچند ممکن است به علت عدم وجود علائم بر حروف، خوانایی کلمه دچار مشکل شود[۱].

۷.۳ تغییر در شکل ظاهری حروف

در این روش با تغییر شیب برخی حروف می‌توان بیت‌های واترمارک را درج کرد. بدین صورت که برای درج بیت یک، شیب برخی حروف مانند حرف "ر" تغییر یافته و برای درج بیت صفر شیب حروف تغییری نمی‌کند. محاسبه شیب یک حرفی مانند حرف "ر" برابر است با $\tan(R)$ که از حاصل تقسیم طول عمودی بر طول افقی حرف "ر"، بدست می‌آید. در شکل ۴ Lr و Wr که معرف طول عمودی و طول افقی ومکان هستند و زاویه R ، نشان داده شده است[۱۳].

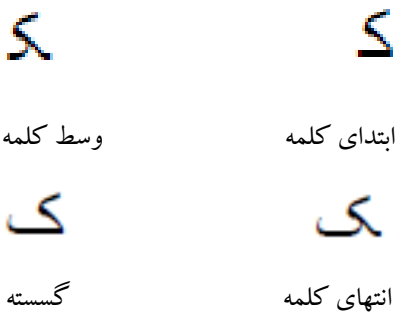


شکل ۴- نحوه تعریف شیب برای حرف "ر" [۱۳]

۴. روش پیشنهادی

۱.۴ مشخصات حروف "ک" و "گ"

حروف فارسی را می‌توان به گروه‌هایی تقسیم کرد که در هر گروه بدنه اصلی حروف یکسان بود و فقط در تعداد نقاط بالا یا پایین، علائمی مانند سرکش و دسته و یا شیب و انحنای حروف تفاوت دارند. حروف "ک" و "گ" از دسته حروف بدون نقطه محسوب شده و ویژگی بارز آن‌ها وجود سرکش است. تنها تفاوت حرف"گ" با حرف "ک" وجود یک سرکش کوچک بر روی حرف "گ" می‌باشد. حروف "ک" و "گ" با توجه به مکان قرارگیری‌شان در کلمه به چهار صورت متفاوت نمایش داده می‌شوند در شکل ۵، اشکال مختلف حرف "ک" نشان داده شده است.



شکل۵- اشکال مختلف حرف "ک" با توجه به مکان قرارگیری حرف در کلمه

همچنین با توجه به نوع فونت، شکل ظاهری این حروف تغییر خواهد کرد. در شکل ۶، تاثیرگذاری چند فونت مختلف بر روی حرف "ک" در کلمه "کاکتوس" نشان داده شده است.

فونت B Nazanin فونت B Sina

فونت B Arshia فونت B Farnaz

شکل۶- اشکال مختلف حرف "ک" باتوجه به نوع فونت

با توجه به آنکه فونت B Nazanin به طور پرکاربرد در اکثر مقالات، پایان نامه ها و... استفاده می شود، لذا تمرکز بر روی این فونت خواهد

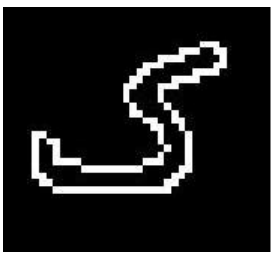
۲.۴. جای سازی اطلاعات

جهت تعبیه اطلاعات ابتدا تصویر متنی که حاوی متون فارسی باشد، به عنوان متن پوششی انتخاب می‌گردد. از آنجایی پیکسل-های تصویر ورودی دارای مقادیر متفاوت از صفر تا ۲۵۵ هستند، لذا از یک آستانه معین جهت تبدیل تصویر ورودی به تصویر سیاه و سفید تک بیتی استفاده می‌کنیم. شکل ۷ تصویر حرف "ک" قبل و بعد از گرفتن آستانه را نشان می‌دهد.

الف) تصویر حرف "ک" قبل از گرفتن آستانه **ب) تصویر حرف "ک" بعد از گرفتن آستانه**

شکل۷- تصویر حرف "ک" قبل و بعد از گرفتن آستانه

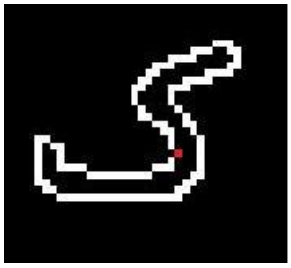
در مرحله بعد از یکی از روش‌های لبه‌یابی موجود در پردازش تصویر جهت استخراج لبه‌های موجود در تصویر متنی باینری استفاده می‌شود. در شکل ۸ لبه‌های موجود در حرف "ک" نشان داده شده است.



شکل ۸ – لبه‌های موجود در حرف "ک"

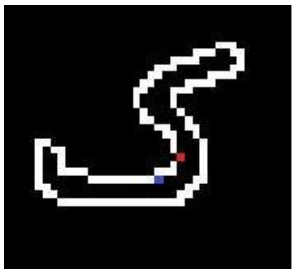
مختصات نقاطی که دارای لبه هستند، در ماتریسی ذخیره می‌شود که به آن ماتریس لبه می‌گویند. سپس ماتریس صفر که هم-اندازه با ماتریس لبه باشد، ساخته می‌شود. به ازای هر پیکسل موجود در لبه، پنجره‌ای با محوریت هر پیکسل و به ابعاد ۳ در۳ تشکیل شده که از هر طرف، به اندازه یک پیکسل ضخامت دارد. در نهایت برای تشکیل ماتریس زاویه، زاویه هر پنجره با محوریت هر پیکسل تعیین می‌شود و مقادیر زوایای بدست آمده در ماتریس صفر جای‌گذاری خواهد شد.

بر اساس ماتریس زاویه بدست آمده، مختصاتی از لبه حرف "ک" یا "گ" که دارای ماتریس ۳ در ۳ با مقادیر زاویه منحصر بفرد در ماتریس زاویه هستند، انتخاب می‌شود. به عنوان نمونه شکل ۹ مکان نقطه مورد نظر را با رنگ قرمز نشان می‌دهد.



شکل ۹- مختصات پیکسل مورد نظر

اکنون بر اساس نقطه بدست‌آمده می‌توان اطلاعاتی که تبدیل به صفر و یک شده‌اند را در تصویر متنی تعبیه کرد. بدین صورت که برای تعبیه بیت با مقدار صفر یک پیکسل با مقدار یک روی لبه حروف "ک" یا "گ" قرار می‌گیرد و برای بیت با مقدار یک نیز همین روند تکرار می‌شود. لازم به ذکر است که پیکسل‌های اضافه شده روی لبه برای بیت با مقدار صفر دارای مختصات طولی و عرضی متفاوتی نسبت به پیکسل‌های تعبیه شده برای بیت یک هستند. شکل ۱۰ تصویر حرف "ک" حاوی بیت صفر و بیت یک را نشان می‌دهد.

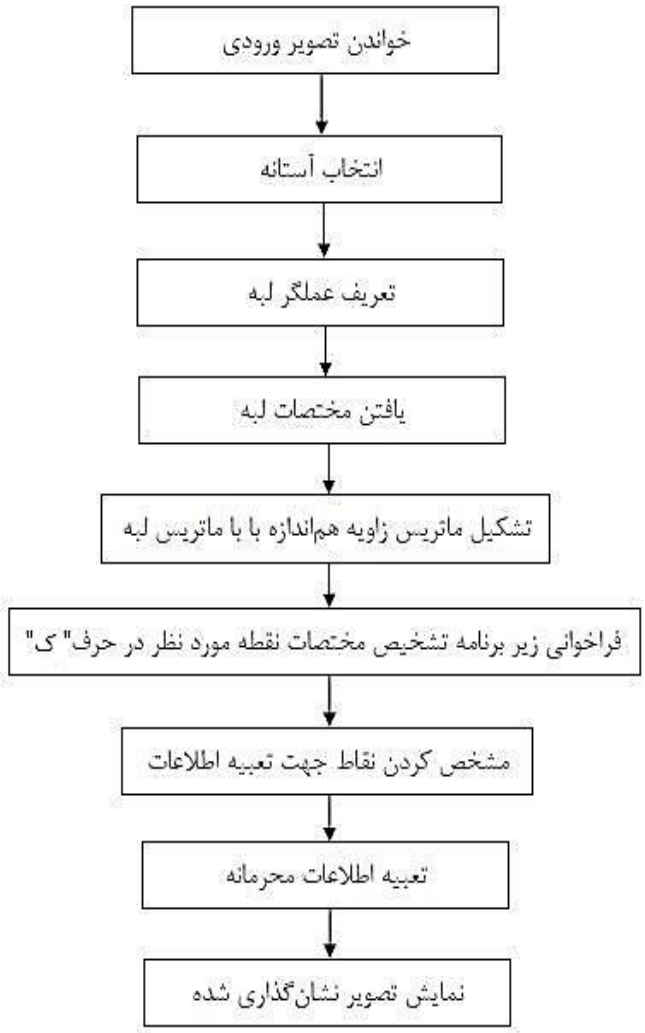


الف) تصویر حرف "ک" حاوی بیت صفر

ب) تصویر حرف "ک" حاوی بیت یک

شکل ۱۰- تصویر حرف "ک" حاوی اطلاعات تعبیه شده

حال برای تعبیه اطلاعات محرمانه، ابتدا اطلاعات را به حالت باینری تبدیل کرده و سپس طبق مراحل شرح داده شده در بالا، با توجه به ارزش بیت محرمانه پیکسل‌های موردنظر را روی لبه حروف "ک" و "گ" تعبیه می‌شود. مراحل تعبیه اطلاعات طبق روش پیشنهادی در شکل ۱۱ نشان داده شده است.



شکل ۱۱- مراحل تعبیه اطلاعات طبق روش پیشنهادی

۳.۴. استخراج اطلاعات

روند استخراج اطلاعات دقیقا عکس روند درج اطلاعات است.برای استخراج بیت‌های واترمارک ابتدا بر اساس ماتریس زاویه بدست آمده، مختصاتی از لبه حرف "ک" یا "گ" که دارای ماتریس ۳ در ۳ با مقادیر زاویه منحصر بفرد در ماتریس زاویه هستند، مشخص می‌شود. سپس بر اساس مختصات نقطه بدست‌آمده می‌توان تشخیص داد که در حروف "ک" و "گ" چه اطلاعاتی تعبیه شده است. با حرکت طولی و عرضی از مختصات نقطه مورد نظر، مکان پیکسلی که حاوی بیت صفر باشد، بررسی می‌گردد. اگر پیکسل دارای مقدار یک باشد، نشانگر این است که بیت صفر تعبیه شده است. همین روند نیز جهت تشخیص بیت یک تکرار می‌شود.با کنارهم قرار دادن بیت‌های صفر و یک، اطلاعات تعبیه شده استخراج می‌گردد.

۵. نتیجه‌گیری

روش پیشنهادی که جزء زیر مجموعه روش های کور می باشد، از عدم مشاهده پذیری بالاتری نسبت به روش های قبلی برخوردار بوده، زیرا تغییراتی نامحسوس در سرکش حروف "ک" و "گ" آن هم در لابه لای تعداد بسیار زیاد کلمات و حروف برای چشم انسان کاملا ناپیداست و گیرنده قانونی متن واترمارک شده تنها با در اختیار داشتن الگوریتم واترمارکینگ مناسب قادر به استخراج خواهد بود. شکل ۱۲ نمونه‌ای از متن اصلی و متن واترمارک شده را نشان می‌دهد.

حروف فارسی را می‌توان به گروه‌هایی تقسیم کرد که در هر گروه بدنه اصلی حروف یکسان بود و فقط در تعداد نقاط بالا یا پایین، علائمی مانند سرکش و دسته و یا شیب و انحنای حروف تفاوت دارند. حروف "ک" و "گ" از دسته حروف بدون نقطه محسوب شده و ویژگی بارز آن‌ها وجود سرکش است. تنها تفاوت حرف"گ" با حرف "ک" وجود یک سرکش کوچک بر روی حرف "گ" می‌باشد

الف) تصویر متن اصلی

حروف فارسی را می‌توان به گروه‌هایی تقسیم کرد که در هر گروه بدنه اصلی حروف یکسان بود و فقط در تعداد نقاط بالا یا پایین، علائمی مانند سرکش و دسته و یا شیب و انحنای حروف تفاوت دارند. حروف "ک" و "گ" از دسته حروف بدون نقطه محسوب شده و ویژگی بارز آن‌ها وجود سرکش است. تنها تفاوت حرف"گ" با حرف "ک" وجود یک سرکش کوچک بر روی حرف "گ" می‌باشد

ب) تصویر متن واترمارک‌شده

شکل ۱۲- نمونه‌ای از تصویر متنی اصلی و واترمارک شده

بدلیل عدم وجود یک برنامه OCR قوی برای زبان فارسی متن پرینت شده به آسانی قابل تبدیل به متن ساده نبوده و لذا تخریب اطلاعات مخفی شده بسیار مشکل خواهد بود. در ضمن اطلاعات مخفی شده کاملا نسبت به کوچک یا بزرگ کردن سایز متن مقاوم بوده و از بین نخواهد رفت. از آنجایی که در این روش از کلید برای تعیین سطرهای خاصی از متن به منظور معکوس کردن روند درج الگوریتم استفاده شده است، امنیت این روش نسبت به سایر روش‌ها بهتر است.برای بهبود امنیت این روش می‌توان از یک کلید استفاده نمود که نشان دهنده شماره سطر متون است. تا در آن سطرها روند جاسازی بیت‌های واترمارک معکوس شود.

از طرفی بدلیل شباهت حروف متون فارسی با متون عربی و اردو این روش برای زبان های دیگر نیز قابل پیاده سازی خواهد بود.این روش نسبت به retyping مقاوم نبوده و دارای ظرفیت بسیار بالا نبوده است. مخصوصا برای متون عربی که فاقد حرف "گ" می‌باشند.

۱۲. مراجع

1. Bensaad, M.L. and Yagoubi, M.B. (2011). *High capacity diacritics-based method for information hiding in Arabic text*, International Conference on Innovations in Information Technology (IIT), vol. 11, no. 12, pp. 433 - 436.

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

3. Jaseena K.U. and Anita John (2011). *Text Watermarking using Combined Image and Text for Authentication and Protection*, International Journal of Computer Applications (0975 – 8887) Volume 20– No.4.

4. Young-Won Kim, Kyung-Ae Moon and Il-Seok Oh (2003). *A text watermarking algorithm based on word classification and inter-word space statistics*, Document Analysis and Recognition, 2003. Proceedings. Seventh International Conference on vol., no., pp. 775- 779.

5. Kiranmayi Penumarthi (2005). *Augmented Watermarking*. Thesis, Master of Science in Electrical Engineering, Louisiana State University and Agricultural and Mechanical College.

6. Shirali-Shahreza, M. (2008). *Pseudo-space Persian/Arabic text steganography*, Computers and Communications, 2008. ISCC 2008. IEEE Symposium on vol., no., pp.864-868.

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

8. M. Hassan Shirali-Shahreza and Mohammad Shirali-Shahreza. (2009). *ARABIC/PERSIAN TEXT STEGANOGRAPHY UTILIZING SIMILAR LETTERS WITH DIFFERENT CODES*, The Arabian Journal for Science and Engineering, Volume 35.

9. Adnan Gutub, Fahd Al-Haidari, Khalid Al-Kahsah and Jameel Hamodi. (2010). *e-Text Watermarking: Utilizing 'Kashida' Extensions in Arabic Language Electronic Writing*, Journal of Emerging Technologies in Web Intelligence (JETWI).

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

11. Jibran Memon, Kamran Khowaja, and Hameedullah Kazi. (2008). *Evaluation of Steganography for Urdu/Arabic* Text, Journal of Theoretical and Applied Information Technology - JATIT, Vol. 4, No. 3, pp. 232-237.

12. Gutub A., Elarian Y. S., Awaideh S. M. and Alvi A. K. (2008). *Arabic Text Steganography Using Multiple Diacritics*, International Workshop on Signal Processing and its Applications, Sharjah, U.A.E.

13. Davarzani, R. and Yaghmaie K. (2009). *Farsi Text Watermarking Based on Character Coding*, International Conference on Signal Processing Systems , vol., no., pp.152-156.