پنجمین کنفرانس انجمن رمز ایران
انجمن رمز ایران
Iranian Society of Cryptology
۲۴ الی ۲۶ مهر ماه ۱۳۸۶ دانشگاه علم و صنعت ایران

ارایه روشی جدید برای کشف اطلاعات پنهان شده در تصاویر

محمد مهدی هاشمی نژاد
دانشگاه تربیت مدرس
hasheminezhad@modares.ac.ir

مهدی اشرفی
دانشگاه تربیت مدرس
m.ashrafi@modares.ac.ir

سالومه توفیقی
toufighi.s@gmail.com

چکیده: پنهان نگاری هنر تبادل محرمانه اطلاعات است و در مقابل آن پنهان گشایی با هدف کشف تبادل محرمانه اطلاعات قرار دارد. یکی از رایج ترین روشهای پنهان نگاری، پنهان سازی اطلاعات در بیت کم ارزش تصاویر نقش بیتی می باشد. در این مقاله با استفاده از الگوریتم های فشرده سازی اطلاعات و روشهای آماری، الگوریتم جدیدی جهت پنهان گشایی اطلاعات پنهان در بیت کم ارزش تصاویر نقش بیتی ارایه می شود. این روش با استفاده از ارتباط بین حجم تصویر فشرده شده و داده های پنهان در تصویر، برای آشکار سازی داده هایی که در تصویر به صورت مجتمع در مکانهای نامشخص قرار داده شده اند، قابل استفاده است.

واژه های کلیدی: پنهان نگاری، پنهان گشایی

۱ – مقدمه

پنهان نگاری[1] هنر تبادل محرمانه اطلاعات است. در این علم، هدف این است که اطلاعات محرمانه (شئ پنهان[2]) در اطلاعات بی اهمیت دیگری (شئ پوشش[3]) مانند یک فایل تصویری، موسیقی، فیلم، متن و ... طوری پنهان شود که شئ پوشش حاوی شئ پنهان از شئ پوشش فاقد آن قابل تشخیص نباشد. در مقابل پنهان نگاری، علم پنهان گشایی[4] قرار دارد که هدف آن تشخیص شئ پوشش حاوی شئ پنهان از شئ پوشش فاقد شئ پنهان می باشد.

برای پنهان نگاری در انواع اشیاء پوشش مختلف الگوریتم های متفاوتی مطرح شده است که تمرکز این مقاله بر روی شئ پوشش از نوع تصویر نقش بیتی[5] و الگوریتم پنهان نگاری در بیتهای کم ارزش[6] شئ پوشش می باشد.

وست فلد[7] با روش حمله بصری *[1]*، نشان داد که بیتهای کم ارزش تصاویر برخلاف آنچه تصور می شد، دارای الگوهایی هستند و در ادامه با ابداع روش *PoV* توانست وجود این الگوها را با استفاده از روشهای آماری بررسی کند و اولین روش پنهان گشایی عملی را ارایه دهد *[1]*. روش *PoV* هنگامی که محل قرار گیری اطلاعات مشخص باشد (به عنوان مثال به صورت ترتیبی در ابتدا یا انتهای فایل حامل)، نتایج قابل اعتمادی ارایه می کند. ولی اگر محل قرار گیری اطلاعات مشخص نباشد، این روش کارایی چندانی ندارد.

فردریچ[8] روش قدرتمند *RS* را برای کشف داده های پنهان در بیتهای کم ارزش تصاویر را ارایه داد *[2,3]*. این روش با استفاده از ۷ بیت بالای هر بایت، مقدار بیت ۸ را پیش بینی می کند و با

[1] Steganography
[2] Stego object
[3] Cover object
[4] Steganalysis

[5] Bitmap
[6] LSB (Least Significant Bit)
[7] Westfeld.A
[8] Fridrich.J

مقایسه مقدار پیش بینی شده با مقدار واقعی، تغییرات در بیت کم ارزش را تخمین میزند. روش *RS* هنگامی که بیتهای حاوی اطلاعات به صورت تصادفی و پراکنده از کلیه بیتهای تصویر انتخاب شده باشند، نتایج بسیار خوبی ارایه میکند و طبق برآورد آندرو کر[9] خطای این روش ۵٪ تخمین زده شده است*[4]*. پس از آن، دومیترسکو[10] به همراه ارایه روش آنالیز جفتهای نمونه[11]، مدل ریاضی دقیقی نیز از روش *RS* ارایه داد*[5]*.

در این مقاله روش جدیدی برای تشخیص وجود اطلاعات در بیتهای کم ارزش تصاویر نقش بیتی ارایه شده است. روش ارایه شده هنگامی که اطلاعات پنهان به صورت متمرکز در نواحی مختلف تصویر حامل پنهان شده باشند، نتایج بسیار دقیقی ارایه می نماید. برای بررسی الگوریتم از ۱۰۸ تصویر موجود در پایگاه تصاویر *KODAK* استفاده شده است*[6]* و این تصاویر به فرمت *BMP* تبدیل شده اند.

این مقاله به شکل زیر سازماندهی شده است. در قسمت ۲ روش معرفی شده در این مقاله ارایه می شود و پارامترهای روش مطرح شده ارزیابی و اعتبار سنجی می شوند. در قسمت ۳ روش مطرح شده آزمایش می شود و نتایج عملی روش ارایه می شود. در نهایت در قسمت ۴ نتیجه گیری انجام می شود.

۲ – معرفی روش پنهان گشایی

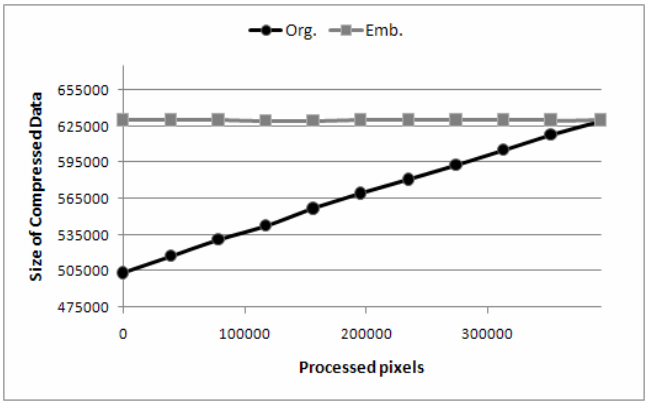
ایده کلی این روش بر این اساس است که پیکسلهای تصویر اصلی دارای نظم خاصی هستند و پیکسلها بعد از تغییر بیتهای کم ارزش تصویر بر اثر اضافه کردن اطلاعات پنهان به نسبت پیکسلها در تصویر اولیه دارای بی نظمی بیشتری خواهند بود.

برای مقایسه میزان بی نظمی حاصل از تغییر بیتهای کم ارزش در پیکسلها روشهای مختلفی میتوان مطرح کرد. در روش ارایه شده در این مقاله، برای بررسی میزان نظم پیکسلها از الگوریتم فشرده سازی *Deflate* استفاده شده است که یک الگوریم فشرده سازی بدون اتلاف[12] و بر اساس ترکیبی از دو الگوریتم *Huffman* و *LZ77* است*[7]*. این الگوریتم در قدم اول رشته‌های تکراری را یافته و این تکرارها را با ارجاع به رشته قبلی جایگزین می‌کند. در قدم بعدی، سمبل‌های ساخته‌شده را توسط کد *Huffman* بهینه‌سازی می‌کند.

انتظار می رود بعد از تغییرات در بیتهای کم ارزش تصویر و افزایش بی نظمی و کاهش رشته‌های تکراری در حامل، بر اساس الگوریتم *Deflate* امکان فشرده سازی اطلاعات کمتر شده و حجم اطلاعات فشرده شده بیشتر شود. برای نمایش این مطلب، تصویر *0795IMAGE01* انتخاب*[6]* و در دو حالت آنالیز شده است.

در حالت اول در تصویر اصلی به ترتیب در فواصل *[0% 10%]* و *[0% 20%]* و ... و *[0% 100%]* داده های تصادفی در بیت کم ارزش قرار داده شد و تصویر در هر مرحله فشرده شد و نمودار تغییرات حجم تصویر در هر مرحله رسم شد*[*شکل ۱: الف*]*. در این نمودار، محور افقی تعداد پیکسلهایی است که در بیت کم ارزش آنها داده تصادفی قرار گرفته و محور عمودی، حجم کل اطلاعات پس از فشرده سازی توسط الگوریتم *Deflate* می باشد.

در حالت دوم، ابتدا در تمامی بیتهای کم ارزش تصویر، داده های تصادفی قرار داده شد و سپس نمودار مشابه حالت اول رسم شد*[*شکل ۱: ب*]*.



شکل ۱: نتایج آنالیز تصویر *0795IMAGE01*

همانطور که انتظار میرفت، شیب خط حالت اول بسیار بیشتر از شیب خط حالت دوم است و شیب خط در حالت دوم نزدیک به صفر است.

[9] Ker.A
[10] Dumitrescu.S
[11] Sample Pair Analysis
[12] Lossless

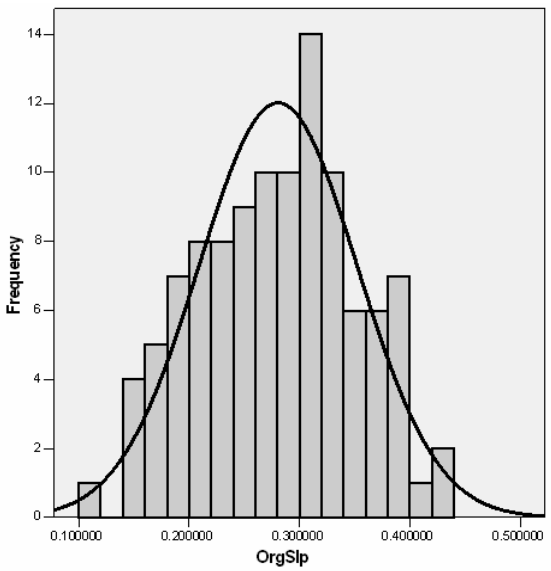
۲-۱ الگوریتم CBS[13]

الگوریتم CBS به شکل زیر پیشنهاد می شود: در تصویر در بازه های $\left[0 \quad \frac{n}{k}\right]$ و $\left[\frac{n}{k} \quad \frac{2n}{k}\right]$ و ... و $\left[\frac{(k-1)n}{k} \quad n\right]$ که در آنها n همان سایز تصویر است، در بیت کم ارزش داده تصادفی قرار داده شود و شیب خط در هر بازه محاسبه شود. میتوان از شیب این خطوط به عنوان سنجه ای برای تشخیص وجود داده پنهان در تصویر استفاده کرد. اگر شیب خط بازه ای از حد آستانه کمتر باشد، در آن ناحیه داده پنهان وجود دارد و اگر شیب خط در بازه ای از حد آستانه بیشتر باشد، در آن ناحیه داده پنهان وجود ندارد.

۲-۲ تعیین پارامترها

در این بخش حدود آستانه برای تشخیص وجود یا عدم وجود داده پنهان با در نظر گرفتن شیب خط حاصل از الگوریتم CBS تعیین می شود. در این قسمت، پارامتر k همواره مقدار $k=10$ در نظر گرفته شده است.

برای تعیین حد بالای آستانه تشخیص عدم وجود داده پنهان در تصاویر الگوریتم CBS بر روی تصاویر اجرا شد و خط رگرسیون نمودار در هر مورد محاسبه شد. در شکل ۲ هیستوگرام و آمار توصیفی مربوط به مقادیر شیب خط رگرسیون ۱۰۸ تصویر مشخص شده است.



Desciptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
OrgSlp	108	.113601	.439786	.2805639	.071720370
Valid N (listwise)	108				

شکل ۲ : هیستوگرام شیب خطوط رگرسیون در تصاویر خام

همانطور که از شکل مشخص است؛ به نظر می رسد که شیب این خطوط از تابع توزیع نرمال پیروی می کند. در شکل ۳ خلاصه تست کلوموگروف-اسمیرنوف برای این توزیع آمده است. همانطور که مشخص است به خاطر مقدار $P\text{-}value = .859$ دلیلی بر رد فرضیه $H0$ مبنی بر نرمال بودن تابع توزیع شیب خطوط نداریم پس می پذیریم که تابع توزیع شیب خطوط رگرسیون حاصل از اجرای الگوریتم CBS بر روی این تصاویر نرمال است.

One-Sample Kolmogorov-Smirnov Test

		OrgSlp
N		108
Normal Parameters[a,b]	Mean	.2805639
	Std. Deviation	********
Most Extreme Differences	Absolute	.058
	Positive	.058
	Negative	-.052
Kolmogorov-Smirnov Z		.604
Asymp. Sig. (2-tailed)		.859

a. Test distribution is Normal.

b. Calculated from data.

شکل ۳: خلاصه تست کلوموگروف-اسمیرنوف بر روی تابع توزیع شیب خطوط رگرسیون در تصاویر خام

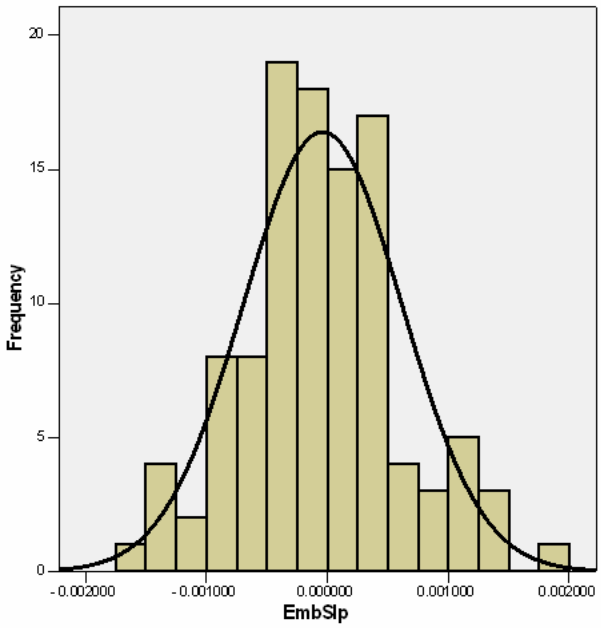
با پذیرفتن فرض نرمال بودن تابع توزیع شیب خطوط، اگر فاصله ۳ سیگما را برای محدوده قابل قبول شیب خط در نظر بگیریم مقدار خطای نوع اول برابر ۰،۰۰۲۷ خواهد بود. در نتیجه بازه $B>0.0654$ را به عنوان حد بالای قابل قبول برای تشخیص اینکه داده ای در تصویر پنهان نشده است در نظر می گیریم و چنانچه شیب خط در هر ناحیه ای از تصویر کمتر از این مقدار شد می پذیریم که در آن ناحیه از تصویر داده پنهان وجود دارد.

برای تعیین حد پایین آستانه تشخیص عدم وجود داده پنهان در تصاویر، ابتدا تمامی بیتهای کم ارزش تصاویر داده تصادفی قرار گرفته شد و سپس الگوریتم CBS بر روی تصاویر اجرا شد و باز هم خط رگرسیون نمودار در هر مورد محاسبه شد. در شکل ۴ هیستوگرام و آمار توصیفی مربوط به مقادیر شیب خط رگرسیون ۱۰۸ تصویر پر از داده های پنهان تصادفی مشخص شده است.

با توجه به شکل 4، بازهم به نظر می رسد که شیب این خطوط از تابع توزیع نرمال پیروی می کند. در شکل ۵ نتایج تست

[13] Compress-Based Steganalysis

کلوموگرف – اسمیرنوف برای متغیر شیب خط در این حالت آمده است. همانطور که مشخص است به خاطر مقدار $P\text{-}value = .838$ دلیلی بر رد فرضیه $H0$ مبنی بر نرمال بودن تابع توزیع شیب خطوط نداریم پس می پذیریم که تابع توزیع شیب خطوط رگرسیون حاصل از اجرای الگوریتم CBS بر روی تصاویر با اطلاعات پنهان نیز نرمال است.



Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
EmbSlp	108	-.001684	.001886	-.000040	.000656700
Valid N (listwise)	108				

شکل ۴ : هیستوگرام شیب خطوط رگرسیون در تصاویر حاوی داده های پنهان

One-Sample Kolmogorov-Smirnov Test

		EmbSlp
N		108
Normal Parameters[a,b]	Mean	-.000040
	Std. Deviation	********
Most Extreme Differences	Absolute	.060
	Positive	.060
	Negative	-.048
Kolmogorov-Smirnov Z		.619
Asymp. Sig. (2-tailed)		.838

a. Test distribution is Normal.

b. Calculated from data.

شکل ۵: خلاصه تست کلوموگروف-اسمیرنوف بر روی تابع

توزیع شیب خطوط رگرسیون در تصاویر حاوی داده های پنهان با پذیرفتن فرض نرمال بودن تابع توزیع شیب خطوط، اگر فاصله ۳ سیگما را برای محدوده قابل قبول شیب خط در نظر بگیریم مقدار خطای نوع اول برابر ۰،۰۰۲۷ خواهد بود. در نتیجه بازه $B<0.0019$ را به عنوان حد پایین قابل قبول برای تشخیص اینکه در تصویر داده پنهان شده است در نظر می گیریم و چنانچه شیب خط در هر ناحیه ای از تصویر بیشتر از این مقدار شد، می پذیریم که در آن ناحیه از تصویر داده پنهان وجود ندارد.

با توجه به دو بازه به دست آمده در دو آزمایش بالا t به عنوان یک حد آستانه در بازه $[0.0019 \quad 0.0654]$ معین می شود تا بر اساس آن برای وجود یا عدم وجود داده در تصویر تصمیم گیری صورت پذیرد.

۳ – نتایج تجربی

برای آزمایش الگوریتم مطرح شده، تصویر $BRIDGEBLUR68$ و همچنین با توجه به قسمت ۲-۲، پارامتر حد آستانه $t=0.03$ انتخاب شد. در بازه های $[0\% \quad 10\%]$، $[40\% \quad 70\%]$ و $[90\% \quad 100\%]$ بیتهای کم ارزش به صورت تصادفی مقدار دهی شد. نتایج اجرای الگوریتم CBS با مقدار $k=10$ بر روی تصویر ذکر شده، بعد از اضافه کردن اطلاعات پنهان، در جدول ۱ آمده است.

جدول ۱: نتیجه آزمایش الگوریتم

نتیجه	شیب خط	بازه مورد بررسی
$<t$ *	-0.00163	[0% 10%]
$>t$	0.336423	[10% 20%]
$>t$	0.281888	[20% 30%]
$>t$	0.2116095	[30% 40%]
$<t$ *	0.001098	[40% 50%]
$<t$ *	-0.00319	[50% 60%]
$<t$ *	-0.00082	[60% 70%]
$>t$	0.173892	[70% 80%]
$>t$	0.300092	[80% 90%]
$<t$ *	-0.00079	[90% 100%]

همانطور که در جدول ۱ ملاحظه می شود، الگوریتم CBS در تمامی بازه ها، به خوبی عمل کرده و بازه های حاوی داده های پنهان را مشخص ساخته است.

۴ –نتیجه گیری

در این مقاله روش جدیدی برای کشف اطلاعات پنهان شده در بیت کم ارزش تصاویر نقش بیتی ارایه شد. الگوریتم *CBS* مبتنی بر استفاده از الگوریتم فشرده سازی بدون اتلاف *Deflate* و استفاده از روشهای آماری جهت بررسی اثر تغییر بیت کم ارزش در حجم تصویر فشرده شده، بنا نهاده شده است. این الگوریتم در حالتی که اطلاعات پنهان به صورت متمرکز در مکانهای نامشخص از تصویر قرار گرفته باشند، کارایی بسیار خوبی دارد. در ادامه با بررسی ۱۰۸ تصویر مختلف *[6]* الگوریتم ارایه شده ارزیابی شده و پارامتر های لازم جهت استفاده از روش مشخص شد و با بررسی یک نمونه نشان داده شد که الگوریتم *CBS* از دقت بسیار خوبی در تشخیص داده های پنهان برخوردار است.

زمینه های پژوهشی زیر را جهت پژوهش های آینده در نظر گرفته ایم. انتظار داریم با بررسی دقیقتر روشهای فشرده سازی بدون اتلاف دیگر، دقت الگوریتم ارایه شده بهبود داده شود. همچنین به نظر میرسد انتخاب بهینه پارامتر k در روش ارایه شده، نیاز به بررسی بیشتر دارد. با انتخاب kهای مختلف و تعیین حد آستانه در هر حالت، می توان حالت بهینه تعادل را برای k پیدا کرد.

۵ – مراجع

[1] A. Westfeld and A. Pfitzmann, "Attacks on Steganographic Systems," Lecture Notes in Computer Science, vol.1768، Springer-Verlag, Berlin, 2000, pp. 61–75.

[2] J. Fridrich, M. Goljan, and R. Du, "Reliable Detection of LSB Steganography in Grayscale and Color Images ", Proc. ACM, Special Session on Multimedia Security and Watermarking, Ottawa, Canada, October 5, 2001, pp. 27–30.

[3] J. Fridrich, M. Goljan, and R. Du, “Detecting LSB Steganography in Color and Gray-Scale Images”, Magazine of IEEE Multimedia, Special Issue on Security, October-November issue, 2001, pp. 22–28.

[4] A. Ker, “Quantitative Evaluation of Pairs and RS Steganalysis”, Proc. SPIE Electronic Imaging, Security, Steganography, and Watermarking of Multimedia Contents VI, San Jose, January 19–22, pp. 83–97, 2004.

[5] S. Dumitrescu, Wu Xiaolin, and Zhe Wang, “Detection of LSB Steganography via Sample Pair Analysis”, In LNCS vol. 2578, Springer-Verlag, New York, pp. 355–372, 2003.

[6] KODAK database. ftp://ftp.kodak.com/www/images/pcd/.

[7] P. Deutsch. DEFLATE compressed data format specification version 1.3. IETF Request for Comments 1951, May 1996.