

ارایه روشی جدید برای کشف اطلاعات پنهان شده در تصاویر

محمد مهدی هاشمی نژاد

مهدی اشرفی

سالومه توفیقی

دانشگاه تربیت مدرس

دانشگاه تربیت مدرس

hasheminezhad@modares.ac.ir m.ashrafi@modares.ac.ir 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^۱ هنر تبادل محرمانه اطلاعات است. در این علم، هدف این است که اطلاعات محرمانه (شیء پنهان^۲) در اطلاعات بی اهمیت دیگری (شیء پوشش^۳) مانند یک فایل تصویری، موسیقی، فیلم، متن و ... طوری پنهان شود که شیء پوشش حاوی شیء پنهان از شیء پوشش فاقد آن قابل تشخیص نباشد. در مقابل پنهان نگاری، علم پنهان گشایی^۴ قرار دارد که هدف آن تشخیص شیء پوشش حاوی شیء پنهان از شیء پوشش فاقد شیء پنهان می باشد.

برای پنهان نگاری در انواع اشیاء پوشش مختلف الگوریتم های متفاوتی مطرح شده است که تمرکز این مقاله بر روی شیء

پوشش از نوع تصویر نقش بیتی^۵ و الگوریتم پنهان نگاری در بیتهای کم ارزش^۶ شیء پوشش می باشد. وست فلد^۷ 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.

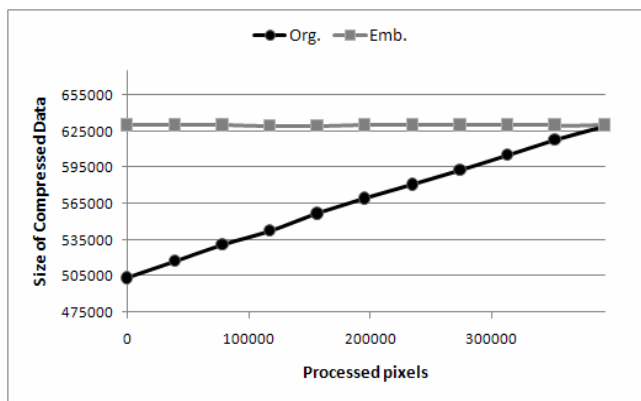
فردریچ^۸ روش قدرتمند RS را برای کشف داده های پنهان در بیتهای کم ارزش تصاویر را ارایه داد [2,3]. این روش با استفاده از ۷ بیت بالای هر بایت، مقدار بیت ۸ را پیش بینی می کند و با

¹ Steganography² Stego object³ Cover object⁴ Steganalysis⁵ Bitmap⁶ LSB (Least Significant Bit)⁷ Westfeld.A⁸ Fridrich.J

قبلی جایگزین می‌کند. در قدم بعدی، سمبل‌های ساخته شده را توسط کد *Huffman* بهینه‌سازی می‌کند. انتظار می‌رود بعد از تغییرات در بیت‌های کم ارزش تصویر و افزایش بی‌نظمی 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 $[0\% 10\% 20\%]$ و $[0\% 100\%]$ 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* می‌باشد.

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.

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 ارائه می‌کند و طبق برآورد آندرو کر^۹ خطای این روش ۵٪ تخمین زده شده است [4]. پس از آن، دومیترسکو^{۱۰} به همراه ارائه روش آنالیز جفت‌های نمونه^{۱۱}، مدل ریاضی دقیقی نیز از روش *RS* ارائه داد [5].

در این مقاله روش جدیدی برای تشخیص وجود اطلاعات در بیت‌های کم ارزش تصاویر نقش بیتی ارائه شده است. روش ارائه شده هنگامی که اطلاعات پنهان به صورت متمرکز در نواحی مختلف تصویر حامل پنهان شده باشند، نتایج بسیار دقیقی ارائه می‌نماید. 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 ارائه می‌شود و پارامترهای روش مطرح شده ارزیابی و اعتبار سنجی می‌شوند. در قسمت ۳ روش مطرح شده آزمایش می‌شود و نتایج عملی روش ارائه می‌شود. در نهایت در قسمت ۴ نتیجه گیری انجام می‌شود.

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 بی‌نظمی بیشتری خواهند بود. برای مقایسه میزان بی‌نظمی حاصل از تغییر بیت‌های کم ارزش در پیکسل‌ها روش‌های مختلفی می‌توان مطرح کرد. در روش ارائه شده در این مقاله، برای بررسی میزان نظم پیکسل‌ها از الگوریتم فشرده سازی *Deflate* استفاده شده است که یک الگوریتم فشرده سازی بدون اتلاف^{۱۲} 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

⁹ Ker.A

¹⁰ Dumitrescu.S

¹¹ Sample Pair Analysis

¹² Lossless

بگیریم مقدار خطای نوع اول برابر 0.027 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 می‌گیریم و چنانچه شیب خط در هر ناحیه ای از تصویر بیشتر از این مقدار شد، می‌پذیریم 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 $[0.0019 \ 0.0654]$ معین می‌شود تا بر اساس آن برای وجود یا عدم وجود داده در تصویر تصمیم‌گیری صورت پذیرد.

۳- نتایج تجربی

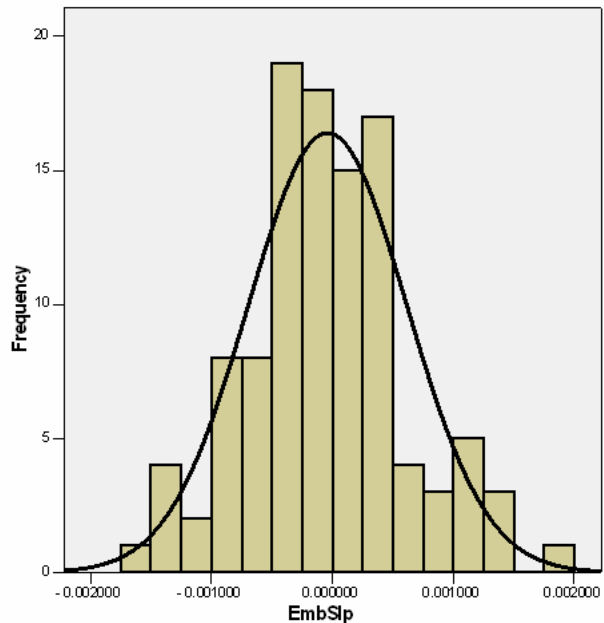
برای آزمایش الگوریتم مطرح شده، تصویر *BRIDGEBLUR68* و همچنین با توجه به قسمت ۲-۲، پارامتر حد آستانه $t=0.03$ انتخاب شد. در بازه های $[0\% \ 10\%]$ ، $[40\% \ 70\%]$ و $[90\% \ 100\%]$ بینهای کم ارزش به صورت تصادفی مقدار دهی شد. نتایج اجرای الگوریتم *CBS* با مقدار $k=10$ بر روی تصویر ذکر شده، بعد از اضافه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

جدول ۱: نتیجه آزمایش الگوریتم

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

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

کلموگروف - اسمیرنوف برای متغیر شیب خط در این حالت آمده است. همانطور که مشخص است به خاطر مقدار $P\text{-value} = 0.838$ دلیلی بر رد فرضیه H_0 مبنی بر نرمال بودن تابع توزیع شیب خطوط نداریم پس می‌پذیریم 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 کلموگروف-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

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 پارامترهای 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 k های 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.

