

نهان نگاری اطلاعات در فایل‌های XML

فرزین یغمایی

دانشکده مهندسی

دانشگاه سمنان

fyaghmaee@semnan.ac.ir

چکیده: مفهوم نهان نگاری^۱ از جمله مفاهیمی است که با گسترش تبادلات دیجیتال و سهولت انتقال اطلاعات اهمیت ویژه ای پیدا کرده است. به طور کلی از نهان نگاری جهت ذخیره سازی اطلاعاتی به صورت پنهان در اطلاعات میزبان^۲ برای اثبات مالکیت و یا فهم تبدیلات و تحریفات در اطلاعات میزبان و یا ارسال اطلاعات به صورت مخفی برای گیرنده خاص استفاده می شود. فایل‌های چند رسانه ای (اعم از صوتی و تصویری) به واسطه حجم زیاد افزونگی^۳ اطلاعات، از قابلیت خوبی جهت نهان نگاری برخوردارند. اما در مقابل فایل‌های متنی از ظرفیت بسیار کمی برای پنهان سازی بهره مندند. با توجه به اهمیت فایل‌های XML در اینترنت و مقبولیت عمومی این ساختار ذخیره سازی اطلاعات، در این مقاله به بررسی ساختار فایل‌های XML پرداخته و سه روش جهت نهان نگاری در آنها ارائه داده ایم که از ظرفیت خوبی در مقایسه با روشهای مرسوم در فایل‌های متنی برخوردارند.

واژه های کلیدی: نهان نگاری، فایل‌های متنی، نهان نگاری در فایل‌های XML.

۱- مقدمه

به طور کلی دو کاربرد عام برای نهان نگاری اطلاعات می توان تصور کرد:

- اطلاعات اضافه و پنهان شده در اطلاعات اصلی حاوی اطلاعات مربوط به اثبات مالکیت است تا بتوان از طریق اطلاعات پنهان شده پی به مالک اصلی اطلاعات برد و یا آنکه بتوان از طریق این اطلاعات فهمید آیا اطلاعات اصلی در کانال ارتباطی موجب تغییر و تحریف شده است یا خیر؟

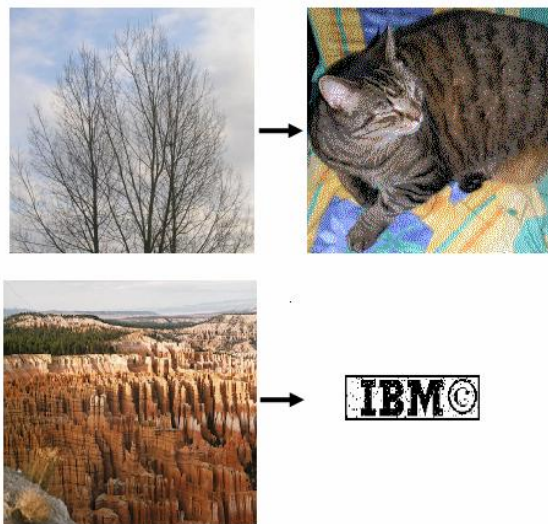
نمایش و انتقال اطلاعات به صورت دیجیتال، عمل انتقال و تکثیر اطلاعات را تسریع و تسهیل کرده و این سهولت به نوبه خود منجر به توزیع غیرقانونی و یا تحریف و تغییر اطلاعات شده است. این مساله سبب شده تا گرایش‌های فراوانی به سمت تکنیک‌های مخفی سازی اطلاعات ایجاد شود تا از طریق این اطلاعات پنهانی بتوان تغییرات و تحریفات را تشخیص داد.

¹ Data hiding

² host

³ Redundancy

در بخش دوم به مقوله نگان نگاری در متن پرداخته و روشهای معمول نگان نگاری در اطلاعات متنی را بررسی می کنیم. در بخش سوم به معرفی فرمت فایل های XML پرداخته و در نهایت در بخش چهارم روشهای پیشنهادی خود در زمینه نگان نگاری در فایل های XML پرداخته و در بخش پنجم به بررسی نتایج می پردازیم.



شکل ۱- مقایسه steganography (تصویر بالا) و الگوگذاری (تصویر پایین) و در فایل های تصویری

۲- نگان نگاری در متن

اطلاعات متنی از جمله مشکلترین انواع اطلاعات جهت نگان نگاری هستند [9]. به عنوان مثال تغییر ۱۰ پیکسل در یک تصویر ۵۱۲x۵۱۲ به هیچ عنوان از دید چشم محسوس نمی باشد در حالی که اضافه کردن یک نقطه یا یک حرف در متن می تواند معنای جمله یا متن را به کل تغییر دهد. این مسأله به علت عدم وجود افزونگی اطلاعات در فایل های متنی است که در فایل های تصویری و صوتی به شدت یافت می شود. دقت کنید منظور ما از فایل های متنی، فایل های با فرمت txt. می باشد و نه تصاویر حاوی متن که به document image معروفند و نگان نگاری آنها در مبحث پردازش تصویر مطرح می شود [3,7,5]. در زمینه نگان نگاری اطلاعات متنی سه روش عمده وجود دارد که به اختصار به آنها اشاره می کنیم.

بدیهی است در این حالت اطلاعات مخفی شده باید در مقابل تبدیلات و حملات احتمالی به اطلاعات اصلی پایدار باقی بماند تا بتوان میزان تغییر و یا مالک اصلی اطلاعات را از روی آن تشخیص داد. انواع روشهای الگوگذاری^۴ نامرئی از این رده محسوب می شوند.

اطلاعات مخفی شده ارتباطی به اطلاعات میزبان نداشته و به تنهایی حاوی اطلاعات مورد نظر بین فرستنده و گیرنده است. به عبارت بهتر بر خلاف حالت قبل که اطلاعات مخفی شده معمولاً حجم کمی داشته و جهت حفاظت اطلاعات میزبان به کار می رود، در این روش اطلاعات ارزشمند همان اطلاعات پنهان شده است که به علت اهمیت موضوع از اطلاعات میزبان جهت پنهان سازی آن استفاده شده است. به این ترتیب در این حالت اطلاعات ارزشمند از نظر گیرنده، همان اطلاعات مخفی شده است و نه اطلاعات میزبان. به همین علت در حالت دوم پایداری در مقابل تغییرات چندان مطرح نیست چرا که در این حالت از اطلاعات مخفی شده تنها فرستنده و گیرنده آگاه هستند.

شکل ۱ مقایسه بین این دو کاربرد را هنگامی که اطلاعات در قالب تصویر هستند نشان می دهد. در تصویر (۱-بالا) تصویر یک گربه با حجم نسبتاً زیاد در تصویر درخت به صورت پنهان وارد شده است. اما در شکل (۱-پایین) یک آرم تجاری با حجم نسبتاً پایین در تصویر قرار دارد. هر یک از دو حالت فوق در کاربردهای خاصی مورد استفاده قرار میگیرند. استفاده از نگان نگاری جهت ارسال اطلاعات به صورت مخفی (حالت دوم) در مباحث پردازش تصاویر به Steganography مشهور است.

به این ترتیب روشهای نگان نگاری براساس حجم اطلاعات مخفی شده، میزان پایداری اطلاعات مقابل تبدیلات، نوع کاربرد و همچنین نوع اطلاعات میزبان (نظیر صوت، تصویر، متن و ..) به انواع مختلفی تقسیم می شوند [9,8,4,1]. اما به عنوان یک اصل کلی هر چه میزان اطلاعاتی که مخفی می شود بیشتر باشد پایداری این اطلاعات در مقابل تبدیلات و تهدیدات کمتر است [9,8]. ادامه مقاله به شکل زیر سازماندهی شده است:

⁴ Watermarking

برای ارزیابی میزان تقریبی نمان نگاری در هر یک از سه روش فوق، ما به طور تصادفی ۵۰۰ صفحه XML با حجم تقریبی بین یک یا دو صفحه را انتخاب کرده و میزان ظرفیت پنهان سازی در هر یک از سه روش ذکر شده را مقایسه کرده ایم. جدول ۱ مقایسه مقدار متوسط بیت ذخیره شده در هریک از روشهای فوق است.

جدول ۱-مقایسه ظرفیت در سه روش پیشنهادی

روش سوم	روش دوم	روش اول	نوع روش
۲۰	۹۴	۴۵	متوسط بیت پنهان شده در صفحه

جدول فوق نشان می دهد روشهای اول و دوم از نظر ظرفیت برتری قابل ملاحظه ای نسبت به روش سوم دارند. به نظر می رسد از روش سوم با توجه به شرایط خاص آن، می توان برای الگوگذاری جهت اثبات مالکیت استفاده کرد.

۶- نتیجه گیری

در این مقاله پس از مرور کوتاهی بر مفاهیم نمان نگاری به معرفی ساختار کلی فایل های XML پرداخته و سه روش جهت پنهان سازی اطلاعات بر اساس مشخصات این زبان معرفی کردیم. به کمک روشهای فوق و یا ترکیبی از آنها می توان تقریباً در هر صفحه وب حدود ۱۵ کارکتر را ذخیره کرد که در مقایسه با روشهای متنی بسیار قابل توجه است. کارهای بعدی در این زمینه می تواند در زمینه یافتن خصایص مناسب در دیگر در فایل های متنی ساخت یافته باشد.

۷- مراجع

1. H. Berghel, "Hiding Data, Forensics, and Anti-Forensics", COMMUNICATIONS OF THE ACM Journal, April 2007/Vol. 50, No. 4
2. S. Zhong1, X. Cheng1,2, T. Chen, "Data Hiding in a Kind of PDF Texts for Secret", Communication, International Journal of Network Security, Vol.4, No.1, PP.17-26, Jan. 2007

با سویچ کردن بین این دو وضعیت می توان بیت های صفر یا یک را مخفی کرد. مثلاً در متن زیر دنباله 01010 مخفی شده است.

```
</img>

</img>

</img>
```

روش ۲: جای خالی در دنباله های کنترلی

ایده این روش که در فایل های متنی ساده نیز وجود دارد استفاده از کارکترهای جای خالی در مواقعی است که در مفهوم جمله چندان تأثیری ندارند. در XML وجود یک جاهای خالی در ابتدای یک دنباله کنترلی تأثیر ندارد. به این ترتیب از کد گذاری زیر برای مخفی سازی 0 و 1 استفاده می کنیم.

```
<tag>, </tag>, or <tag/> (hidden) →0
<tag >, </tag >, or <tag /> (hidden) →1
```

متن زیر اطلاعات 101100 010011 را ذخیره کرده است.

```
<user ><name>Farzin</name ><id>01</id></user>
<user ><name>Thirdparty</name><id>02</id ></user >
```

روش ۳: جابجایی در ترتیب دنباله کنترلی:

یک روش دیگر برای پنهان کردن اطلاعات، جابجایی در ترتیب دنباله کنترلی عناصر است به صورتی که نحوه نمایش از دید کاربر تغییر نکند. البته این روش برای صفحات XML با اطلاعات خاص و ساخت یافته مناسب است. مثلاً در دو دستور زیر جای متغیر Date نسبت به Month مشخص می کند که منظور ذخیره سازی اطلاعات است صفر است یا یک؟

```
<event month="MONTH"
date="DATE">EVENT</event> (hidden) →0
<event date="DATE"
month="MONTH">EVENT</event> (hidden) →1
```

در متن زیر دنباله 01 ذخیره شده است.

```
<event month="MEHR" date="1">Ceremony
day</event>
<event date="15" month="AZAR">Birthday </event>
```

ضمناً با توجه به آنکه روشهای فوق از یکدیگر مستقل هستند، امکان ترکیب روشهای فوق و بالا بردن ظرفیت نیز وجود دارد.

۵-مقایسه ظرفیت در روشهای فوق



3. A.Mikkilineni, P.Chiang, E.J. Delp," Data Hiding Capacity and Embedding Techniques for Printed Text Documents", Proceedings of the IS&T's NIP22, International Conference on Digital Printing Technologies, Denver, CO, September 17, 2006, pp. 444-447
4. C.Roberto, G. Ryouiske,"Data Hiding in Identification and Offset IP fields", Fifth International. Symposium on Advanced Distributed Systems,IEEE ISSADS 2005
5. A.Mikkilineni, G. Ali, J. P. Allebach, E. J. Delp, "Signature-embedding in printed documents for security and forensic applications", Proceedings of the SPIE International Conference on Security, Steganography, and Watermarking of Multimedia Contents VI, Volume 5306, San Jose, CA, January 2004, pp. 455-466
6. M.Levene, P.Wood, "XML Structure Compression", Technical Report BBKCS-02-05, School of Computer Science and Information Systems, Birkbeck College, University of London, 2002
7. Q. Wong, N. Memon, "Data Hiding in Binary Text Documents", Security and Watermarking of Multimedia Contents, San Jose, CA, February 2001
8. F. Petitcolas, R.Anderson, M.Kuhn , "Information Hiding,A Survey", , Proceedings of the IEEE, special issue on protection of multimedia content, July 1999.
9. W. Bender, "Techniques for data hiding", IBM SYSTEMS JOURNAL, VOL 35, NO 3&4, 1996
10. M. Shirali-Shahreza, "A New Method for Steganography in HTML Files," Proceedings of the International Joint Conference on Computer, Information, and Systems Sciences, and Engineering (CISSE 2005), Bridgeport,December 2005