http://ansinet.com/itj

ITJ

ISSN 1812-5638

INFORMATION TECHNOLOGY JOURNAL

ANSInet
Asian Network for Scientific Information
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

Information Technology Journal 10 (4): 889-893, 2011
ISSN 1812-5638 / DOI: 10.3923/itj.2011.889.893

Steganography in Ms Excel Document using Text-rotation Technique

[1,2]Bin Yang, [2,3]Xingming Sun, [2]Lingyun Xiang, [2]Zhiqiang Ruan and [1]Ruizhen Wu
[1]Department of Information Engineering and Technology, NanHai Campus,
South China Normal University, Foshan, 528200, China
[2]College of Information Science and Engineering, Hunan University, Changsha, 410082, China
[3]College of Computer and Software, Nanjing University of Information Science and Technology,
Nanjing, 210044, China

Abstract: Most text steganographic methods are taken the formatted text documents, such as MS Word, PDF, PPT and etc., as cover carriers to hide secret information. This study concerns on the steganography in MS Excel document and proposes a new steganographic method hiding information efficiently by text-rotation technique. The proposed method is implemented by slightly rotating the angle of the text inside the cell to reduce the visible detection of the embedded information. Measuring the text angle of the cells retrieves the secret information. Experiments for different threshold in the algorithm are presented and the results show the proposed method not only has a good imperceptibility but also achieve high embedding rate while most of cells in Excel document are short in length.

Key words: Steganography, MS excel document, text rotating, text steganographic method, human visual system

INTRODUCTION

Steganography is concerned with hiding information in redundant space of any unremarkable cover medium and keeps the secret information undetectable without destroying the cover medium integrity (Gutub and Fattani, 2007).

As a technique of protecting the secret information, steganography does not focus on limiting or controlling the access to information, but protecting the hidden information not to be detected or destroyed. Instead of how the encryption protects data, steganography hides the very existence of the information (Provos and Honeyman, 2003). Capacity, security and robustness, which are the three main factors that influence steganography, are contending with each other. Capacity regards to the number of secret information bits could be hidden in the cover medium. Security refers to the possibilities to figuring out the hidden informationby the eavesdropper. Robustness relates to the amount of modification the stego-medium can withstand before the adversary destroys the hidden information (Chen and Wornell, 2001). It should be seeking for the appropriate balance between the three aspects according to the specific requirement.

In the past decade, a number of steganographic methods have been proposed, however, the majority of them use cover medium such as pictures (Chandramouli and Memon, 2001), video clips (Doerr and Dugelay, 2003) and sounds (Gopalan, 2003). In spite of that, text documents are the most prevalent and indispensable form of information nowadays and always be used as a cover medium. For instance, Margaret Thatcher, former British Prime Minister used to put certain number of white-spaces in documents associated with each minister to identify the owner of the document in order to prevent disclosure of government documents by the press (Moerland, 2003).

Most text steganography are based on the formats of TXT, MS Word, PDF, PPT, etc. However, few works have studied on steganographic methods of embedding data in MS Excel document. This paper considers the problem of steganography in MS Excel document for the purpose of hiding additional information in it.

Many works have been done on text steganography. Since mostly Excel documents have little natural language text, the linguistic steganographic schemes (Murphy and Vogel, 2007; Bennett, 2004) are inefficient. Hence, we do not consider the linguistic steganography in this study. The list of several different methods, which could be use in Excel documents, are as follow.

Corresponding Author: Xingming Sun, School of Computer and Communication, Hunan University, No.252, Lushan South Road, Yuelu District, Changsha, 410082, China Tel: 86-731-88821341 Fax: 86-731-88821341

Open space method (Brassil *et al.*, 1999): Hiding information is done through embedding information by utilizing the space characters in a plain text document in this method. These space characters are placed at the end of paragraph lines or between the words. Open space method could be used more widely for the Excel document. Most cells in a MS Excel document are empty so that these can be used for embedding space without changing the document's appearance.

Character features methods (Rabah, 2004): Some features of characters in a text are changed to embed information in these methods. For example, it can change the font color slightly to embed the information in the text. Steganography based on character features can hold a large quantity of secret information without making normal readers aware of the existence of such information in the text. Most steganography methods base on character feature are suitable for Excel document, such as changing the text size, color, alignment, etc. However, modifying a single character without affecting the whole string in the cell is notpossible in an Excel document. Therefore, the steganographic effect in an Excel document is not as good as that in the Word document.

Line and word shifting methods (Low *et al.*, 1995): Shifting text lines vertically and shifting words horizontally could embed information in these methods. Security of this method depends on the availability of varying the distances between words and lines to puzzle intruders. For example, some lines are shifted 1/300 inch up or down in the text and information are hidden by creating a hidden unique shape of the text. But these methods can not be directly applied in Excel document. However, we could change the height of rows or the width of columns slightly to embed information.

Abbreviation methods (Bender *et al.*, 1996): Another method for hiding information is using of abbreviations. Usually, there are few words could be translated to abbreviations in a text, so very little information can be hidden in the text, as in the case of Excel document.

File structure methods (Cantrell and Dampier, 2004; Castiglionea *et al.*, 2007): Many documents contain readily available spaces that can be used inside their file structures. In these methods, some unused space is used to embed information. Meta-data is an example; it is ingrained in file structures but not visible to the user without special tools. Some files also have unused space. In these spaces, bits can be overwritten without any adverse or obvious effect on the file. These spaces create an opportunity to hide information. In Excel file, there are 420 bytes continuous block below the header where can embed data easily.

IMPLEMENTATION

Here, steganography scheme in MS Excel document based on text-rotation algorithm is presented. The proposed steganography scheme described below composes of the embedding and extracting process. This study proposes a novel steganographic method by slightly rotating the angle of the text in the cells according to the value of the corresponding secret bit.

The general algorithm: The proposed hiding Algorithm consists of three steps. In the first step, encoder encrypt the secret information by some encrypt methods such as RC2, RC4, DES etc. In the second step, the encoder first scans the Excel document to obtain the text in a cell, then judges whether the currently processing text is embeddable. Here, 'embeddable' means that the length of currently processing text is less than a threshold, and it will be discussed in the next section. In the last step, the text-rotation algorithm is used for embedding the data, and outputs the stego-document. Figure 1 depicts the embedding process of the proposed algorithm. The basic idea in the third step is to make sure attackers cannotdistinguish the difference between the stego-document and the original document with the naked eye.

Text rotates in MS excel document: In general, HVS (Human Visual System) is hard to detect the minor changes in the text's angle. However, being more than 1° rotation, the changes in text is obvious, therefore only 1° rotation is considered. Table 1 shows the comparison between original text and 1° rotated text.

However, the Fig. 1 shows that the stego-text gets worse with the increase of text's length in a cell. After 1° rotation, the appearance of the first three characters almost have no changed and every five characters compose a group in the behind characters. Every character in a group is in the same horizontal line. Figure 2 shows that each group will raise or fall of a point comparing to the former group after rotation.

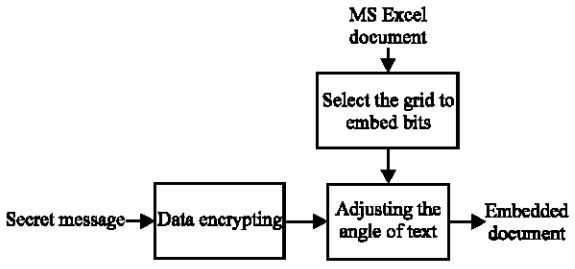

Fig. 1: The embedding process

Table 1: Comparison between original text and 1° rotated text. (The font style is MS reference sans serif and the font size is 14)

Original text	1° rotated text
a	a
bb	bb
ccc	ccc
dddd	dddd
eeeee	eeeee
Continue	Continue
Congratulations	Congratulations
Antidisestablishmentarianism	Antidisestablishmentarianism

Table 2: Comparison between 1°rotated numeric text and-1° rotated numeric text. (The font style is MS Reference Sans Serif and size is 14)

Original	1° rotated	-1° rotated
1	1	1
22	22	22
333	333	333
4444	4444	4444
55555	55555	55555

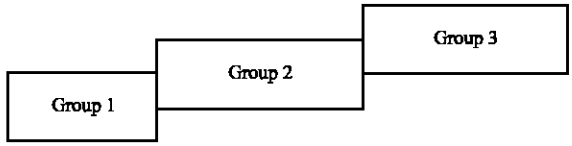

Fig. 2: Groups' relative height after rotation

Since the data type mostly is numeric in the Excel documents, numerical data in the cells should be given more attention. As it was shown in Table 2, while rotating the numeric data by one degree, the alignment of the data will be left-aligned into a right-aligned. Furthermore, because the default alignment of a numeric data is left-aligned, the changes of appearance are magnificent. Nevertheless, if rotating the numeric data by-1°, its alignment remain left-aligned. The data in cells should be treated separately, 1° rotation for the text data and-1° rotation for the numeric ones.

Table 1 and 2 have shown that when the text's length in a cell is less than four, the rotation is hard to detect.

Embedding algorithm based on text-rotationt technique: The proposed algorithm consists of three steps. The first step is finding out the non-empty cell and judge whether its length less than threshold p. The larger the threshold p, the lower the imperceptibility; the imperceptibility is more important than embedding rate in the steganography process, therefore, the threshold p can not be too larger. Threshold p is suggested to be 4 in order to get the good imperceptibility. The second step is to determine the contents of the cell are numeric data or character data. The final step is to changes the angle of cell contents according to the type of data. The details of the message embedding process are presented in the algorithm:

Algorithm 1: Embedding process based on text-rotation technique

Input: MS Excel document, threshold p, secret bits
Output: stego-document
Body:
- For each non-empty cell G do
  - Get the selected cell's length n
  - If n≤p and the secret bit is 1 then
    - If the type of G is text then
      Rotate the angle of G to 1°
    - Else if the type of G is numeric then
      Rotate the angle of G to-1°
- Out put the embedded document

Data extraction process: The extracting process essentially reverses the embedding process. First, the extractor check the cell if it's embeddable. Then, exam the angle and data type from the selected cell to get the secret data. Finally, decrypt the encrypted data and get the secret information.

RESULTS AND DISSCUSION

The proposed methods were implemented using Microsoft Excel 2003 and Microsoft Visual C++6.0 software and running on the Pentium Dual, 2.2 GHz CPU, and 1 GB RAM hardware platform.

In experiments, a student transcript was used as a cover which was shown in Table 3 and the font style of the text was Song and font size was 12. Assuming the secret message bits to be embedded is "10110011000010110010110000110001110000". RC4 was used to encrypt the secret message.

Table 3 compared an original document with the one in which the secret message had been embedded by algorithm1. Note that the difference between the stego-text and the cover text is virtually unnoticeable. In this experiment, the threshold p was 4.

Finally, several configurations for the threshold p of the algorithm1 were used. While changing the threshold p, the embedding rate mostly changed, and the imperceptibility would be changed as well. Figure 3 showed that the embedding rate would increase while the threshold p rose. But setting the threshold too large often leads to poor imperceptibility. Table 4 showed the effect of using large threshold p and the hidden effect was relatively poor.

The embedding rate was closely related to the cover document. Short length text in cells in Excel could result in a high embedding rate. In extreme cases, the embedding

Table 3: Comparison between cover document and stego-document

No.	Mathematics	English	Physics	Chemistry	Biology	Total	Avg.
Cover document							
2010225501	90	63.5	8	92	95	348.5	69.7
2010225502	78.5	77	64	54	73	346.5	69.3
2010225503	5	45	58.5	9	70	187.5	37.5
2010225504	74.5	61	86.5	74	59	355	71
2010225505	72	93.5	64	70.5	93	393	78.6
2010225506	96.5	98	79	80	74	427.5	85.5
2010225507	89	71	73.5	95	71	399.5	79.9
Stego-documnent							
2010225501	90	63.50	8	92	95	348.5	69.7
2010225502	78.5	77	64	54	73	346.5	69.3
2010225503	5	45	58.5	9	70	187.5	37.5
2010225504	74.5	61	86.5	74	59	355	71
2010225505	72	93.5	64	70.5	93	393	78.6
2010225506	96.5	98	79	80	74	427.5	85.5
2010225507	89	71	73.5	95	71	399.5	79.9

Table 4: Effect of setting threshold p to 11, bits '1101001110010110000l' embedded

No.	Mathematics	English	Physics	Chemistry	Biology	Total	Avg
20102205501	90	63.5	8	92	95	348	69.7
20102205502	78.5	77	64	54	73	364.5	69.3

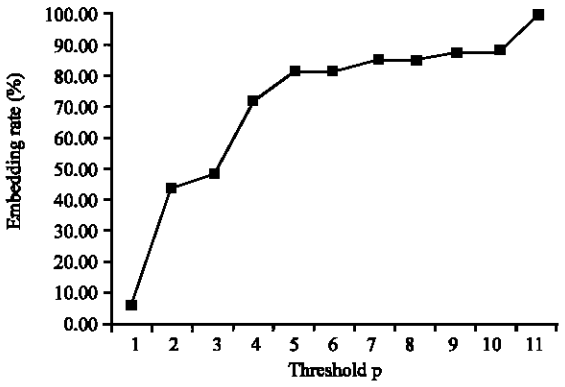

Fig. 3: Embedding rate wiyh difference threshold p

rate would close to 100%, while still maintaining good concealment of the stego-document.

CONCLUSION

The contribution of this study is to develop a novel algorithm for embedding data in an Excel document, and the proposal of it is to use text-rotation technique for steganography. It benefits from the feature that HVS is hard to detect the minor changes in the text's angle. Since many cells in Excel document are very short in length, the embedding rate would be very high.

Although various data hiding methods which based on the text documents are focus on the formats of TXT, MS Word, PDF, PPT, etc., few have studied methods of embedding data in MS Excel document. Compared with other text formats, the expression form of Excel document data are quite different. Therefore, steganography based on Excel document deserves to be further investigations. Many text-based steganographic methods can be used in Excel document, the proposed technique, which is characterized by high capacity and security, will be complementary to the steganographic methods for Excel document.

ACKNOWLEDGMENTS

This study was partially or fully supported by National Basic Research Program of China (973 Program) under Grant No. 2009CB326202 (2009.4-2011.8, Hunan University) and 2010CB334706 (2010.4-2013.3, Hunan University); Key Program of National Natural Science Foundation of China under Grant No. 60736016 (2008.1-2011.12, Hunan University); National Natural Science Foundation of China under Grant No. 60973128 (2010.1-2012.12, Hunan University), 60973113 (2010.1-2012.12, Hunan University), 61073191(2011.1-2013.12, Hunan University) and 61073196 (2011.1-2013.12, Hunan University); Guangdong Province Natural Science Foundation under Grant No. 8151063101000040 (2008.5-2011.4, South China Normal University); the 3rd Guangdong Province 211 program for key subject development (2008.1-2012.12, South China Normal University).

REFERENCES

Bender, W., D. Gruhl, N. Morimoto and A. Lu, 1996. Techniques for data hiding. IBM Syst. J., 35: 313-336.

Bennett, K., 2004. Linguistic steganography: Survey, analysis and robustness concerns for hiding information in text. CERIAS Technical Report, Purdue University, West Lafayette, IN 47907-2086. https://www.cerias.purdue.edu/assets/pdf/bibtex_archive/2004-13.pdf.

Brassil, J.T., S. Low and N.F. Maxemchuk, 1999. Copyright protection for the electronic distribution of text documents. Proceedings of the IEEE, July 1999, IEEE Xplore, London, pp: 1181-1196.

Cantrell, G. and D.D. Dampier, 2004. Experiments in hiding data inside the file structure of common office documents: A steganography application. Proceedings of the International Symposium on Information and Communication Technologies, June 16-18, Las Vegas, Nevada, USA., pp: 146-151.

Castiglionea, A., A.D. Santisa and C. Sorienteb, 2007. Taking advantages of a disadvantage: Digital forensics and steganography using document metadata. J. Syst. Software, 80: 750-764.

Chandramouli, R. and N. Memon, 2001. Analysis of LSB based image steganography techniques. Proceedings of the International Conference on Image Processing, Oct. 7-10, IEEE Computer Society, Washington DC., USA., pp: 1019-1022.

Chen, B. and G.W. Wornell, 2001. Quantization index modulation: A class of provably good methods for digital watermarking and information embedding. IEEE Trans. Inform. Theory, 47: 1423-1443.

Doerr, G. and J.L. Dugelay, 2003. A guide tour of video watermarking. Signal Process. Image Commun., 18: 263-282.

Gopalan, K., 2003. Audio steganography using bit modification. Proceedings of the International Conference on Acoustics, Speech and Signal Processing, April 6-10, IEEE Computer Society, Washington, DC. USA., pp: 412-424.

Gutub, A. and M. Fattani, 2007. A novel arabic text steganography method using letter points and extensions. Proceedings of the WASET International Conference on Computer, Information and Systems Science and Engineering, May 25-27, Vienna, Austria, pp: 28-31.

Low, S.H., N.F. Maxemchuk, J.T. Brassil and L. O'Gorman, 1995. Document marking and identification using both line and word shifting. Proceedings of the 14th Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM'95), April 2-6, IEEE Computer Society, Washington, DC. USA., pp: 853-860.

Moerland, T., 2003. Steganography and Steganalysis. Universiteit Leiden, Rhone-Alpes, France.

Murphy, B. and C. Vogel, 2007. The syntax of concealment: Reliable methods for plain text information hiding. Proceedings of 9th Conference on Security, Steganography and Watermarking of Multimedia Contents, San Jose, CA.

Provos, N. and P. Honeyman, 2003. Hide and seek: An introduction to steganography. IEEE Secur. Privac., 1: 32-44.

Rabah, K., 2004. Steganography-the art of hiding data. Inform. Technol. J., 3: 245-269.