Proposed Data Hiding Technique – Text under Text

Yaman F. Abdullah,[a] Hebah H. O. Nasereddin [b]
[a] MEU—Middle East University, Amman, Jordan
yamanfakhry@hotmail.com
[b] MEU—Middle East University, Amman, Jordan
hebah66@hotmail.com

Abstract. There are many methods used for data hiding like image, audio and text. Data hiding in text is embedding text within another text to be unreadable. Open space methods used for data hiding in text, and white space method is one of these methods; the paper takes advantages of the unused white space from the text "Cover Text" to hide the data "Secret Data" on the cover text with format the secret text by changing the font size and font color.

Keywords: steganography, host signal, signal encoder, white space, syntactic, semantic, Microsoft Word, UniSpaCh, justified format, secret text, cover text.

1 INTRODUCTION

Information security has two branches; data encryption and data hiding. Data encryption is masking the data to become meaningless, While Data hiding is concerned with concealing the data to become unreadable.

(Nasereddin & Al Farzaeai, 2010) Famous examples of steganography go back to antiquity. According to a story from Herodotus, a slave's head was shaved by his master, Histiæus, and tattooed with a secret message around 440 B.C. After growing the hair back, the message disappeared and then the slave journeyed to carry the message. When he shaved his head upon arriving, the message was revealed. In 1860, the major problems had been solved to make a small image by "Darjun", who is a French photographer worked in the war Frank and France in 1870- 1861 when Paris was besieged, by writing messages on photographic films which were sent by the carrier pigeon. The purpose of this was to invoke disobedience against his antagonist Persinas.

Data hiding is:

(Bender, Gruhl, Morimoto, & Lu, 1996) "Data hiding represents a class of processes used to embed data, such as copyright information, into various forms of media such as image, audio, or text with a minimum amount of perceivable degradation to the "host" signal"

1.1 Data Hiding Techniques: (Bender, Gruhl, Morimoto, & Lu, 1996)

1. Data hiding in image: It is a popular technique to hide secret message in image using steganography encode.
2. Data Hiding in Audio: This technique is proposed in many researches to hide secret messages within audio signals using signal encoder.
3. Data Hiding in Text: Some of researchers talk about hiding text within text, this technique can be implemented via three methods; white space on the page, syntactic methods and semantic methods.

1.2 RELATED WORKS

(Brassil, Low, Maxemchuk, & O'Gorman, 1995)

Two types of process are mainly followed to achieve document marking; altering the text formatting, or altering certain characteristics of textual elements. The below mentioned techniques are conveniently used as post-decryption techniques; which enable them to be readable to all:

1. Line-Shift Coding.
2. Word-Shift Coding.
3. Feature Coding.

(Bender, Gruhl, Morimoto, & Lu, 1996)

Discussed data hiding in text. The paper considered the three major open space methods: white space, syntactic methods, and semantic methods for encoding. Also, the paper supported the usefulness of open space methods with two reasons. First, changing the number of trailing spaces has little chance of changing the meaning of a phrase or sentence. Second, a casual reader is unlikely to take notice of slight modifications to white space. The Paper presents three white space methods; encoding a binary message into a text after each terminating character, encoding data by inserting spaces at the end of lines and by encoding data that involves justifying format of text where the extra spaces are placed.

(Por, Wong, & Chee, 2012)

Based on Microsoft Word Document to hide date within text by using DASH to indicate the location of the hidden text, and by using UniSpaCh technique to counter DASH AND analyze it, the Unicode space characters are inserted between words, lines and end of lines. UniSpaCh also used to retrieve the embedded information and reconstruct the original Microsoft Word document.

(Vill´an, et al., 2006)

Proposed a new theoretical framework for data hiding which is concerned with text format feature manipulation to hide text within text by quantizing the color or luminance intensity for each character in a way the human cannot distinguish the modification in the original characters.

(Rahma, AbdulWahab, & Al-Noori, 2011)

Proposed a technique to take advantage of physical file format to store files in the system, the proposed technique used the unused blocks in Microsoft Compound Document File Format (MCDFF). The proposed system embeds text in the file structure (Binary File Format) of document file which is a file of Microsoft Word Document file.

3 METHODOLOGY

3.1 Open space method

Open space method is one of the first used methods to hide data in white space between words, lines and paragraph, this method is divided into three methods;

1. Encodes a binary message into a text by placing either one or two spaces after each terminating character.

2. Encoding data by inserting spaces at the end of lines; two spaces encode one bit per line, four encode two, and eight encode three (See Figure 1).

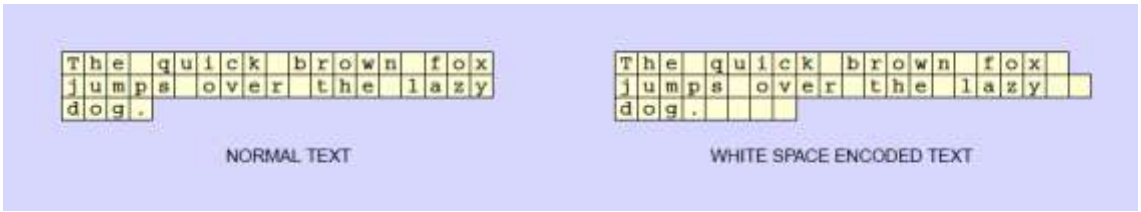

Figure 1: Example of data hiding using white space

3. Encoding binary data by taking advantage of justified format of the text by indicating where the extra space is, and set one space as "0" and two spaces as "1" (See Figure2).

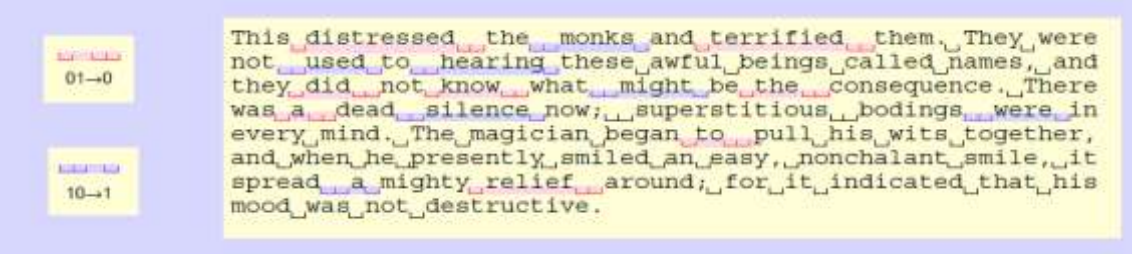

Figure 2: Data hidden through justification (text from A Connecticut Yankee in King Arthur's Court by Mark Twain)

3.2 The problems of open space method are:

- The size of the text output of the secret text hidden within the cover text is very large; where it takes 8 spaces available in the cover text for each character in the secret text. In other words, a very large cover text is required to hide small secret text.
- When encoding data that involves justified format of text. E.g.: to hide two words like "Top Secret" requires text size cover of more than 80 words; because each character size is 8 bit "1 byte" and each bit requires one space.
- In justified format of text, not all spaces are available and can be used to hide data.

4 PROPOSED SOLUTION

4.1 The structure of the proposed model is described as the following:

1. Changing the nature of the secret text: To format the secret text by changing the font size to 1px and set the font color to white as the back color of cover text (See Figure3).

2. Merge secret text within covered text: After checking the secret and covered texts, it will merge the secret text within the covered text using Enhanced Open Space Method; by setting some of characters between words depending on the font size of the cover text like: set three characters in the space if font size if cover text is 12px (See Figure3).

3. Hide the text: At last; it will generate the final text and export it to be transmitted by any e-media like: Internet, CD's, Flash Disk's, and emails (See Figure3).

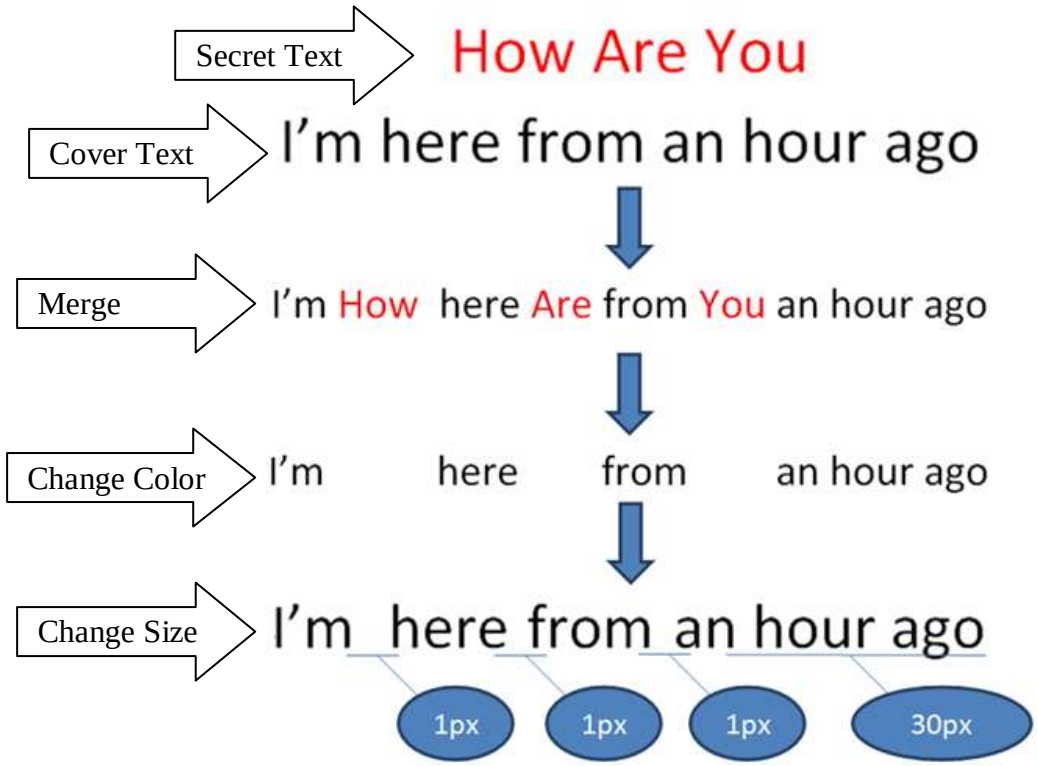

Figure 3: Proposed Data Hiding Technique – Text under Text

4.2 The advantage of proposed solution:

- Size of cover large suitable.
- Solved symbols problem.
- Solved capital and small letter characters.
- Used from all available spaces.
- The output text similar of cover text.

5 RESULTS

5.1 Hide Data

See figure 4; it's the result of hide data by using the proposed solution of data hiding technique – text under text

Figure 4: The result of hide data by using the proposed Data Hiding Technique – Text under Text

5.2 Retrieve Data

See figure 5; it's retrieve the hidden data by using the proposed solution of data hiding technique – text under text

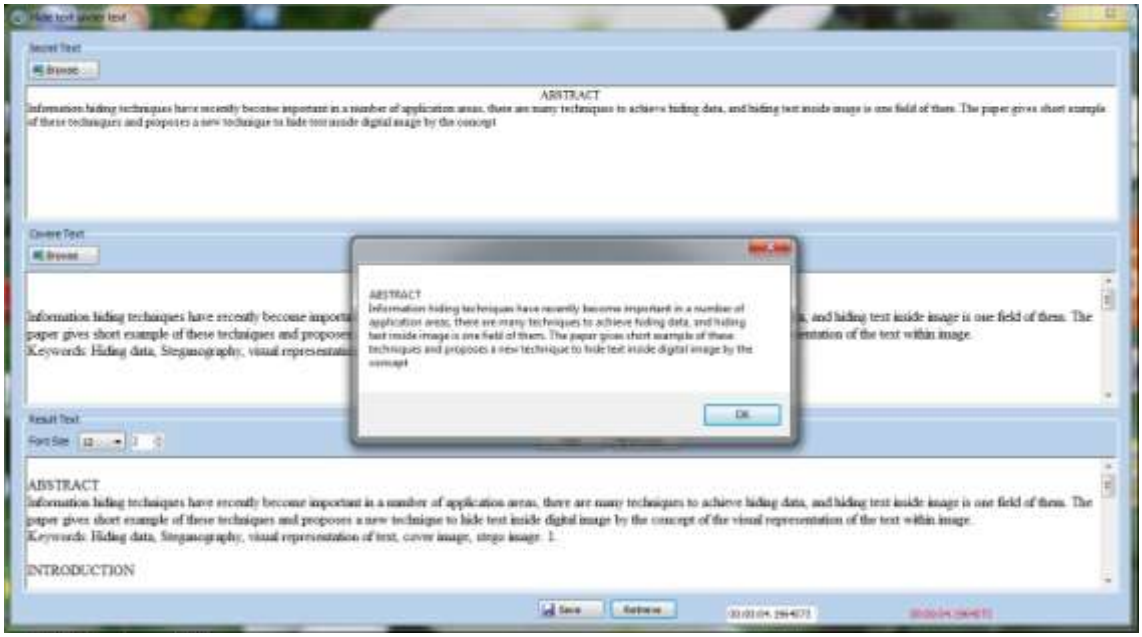

Figure 5: retrieve the hidden data using the proposed Data Hiding Technique – Text under Text

6 CONCLUSION

Image, audio, and text are used for data hiding; Data hiding in text is to embed text within another text to be unreadable. Open space methods used for data hiding in text and white space method is one of these methods; the paper takes advantages of the unused white space from the text "Cover Text" to hide the data "Secret Data" on the cover text. Changing the format of the secret by setting the text size to 1px, setting the font color to white as the back color of cover text, and then merging the secret text with the cover using white space method to generate the result text hiding the secret message within it.

References

Bender, W., Gruhl, D., Morimoto, N., & Lu, A. (1996). Techniques for data hiding. IBM SYSTEMS JOURNAL, 313-336.

Brassil, J. T., Low, S., Maxemchuk, N. F., & O'Gorman, L. (1995). Electronic Marking and Identification Techniques to Discourage Document Copying. Institute of Electrical and Electronics Engineers (IEEE), 1495 – 1504.

Nasereddin, H. H., & Al Farzaeai, M. S. (2010). PROPOSED DATA HIDING TECHNIQUE TEXT IMAGE INSIDE IMAGE (TIII). International Journal of Research and Reviews in Applied Sciences, 183- 193.

Por, L. Y., Wong, K., & Chee, K. O. (2012). A text-based data hiding method using Unicode space characters. The Journal of Systems and Software, 1075 – 1082.

Rahma, A. S., AbdulWahab, H. B., & Al-Noori, A. Y. (2011). Proposed Steganographic Method for Data Hiding in Microsoft Word Documents Structure. Al-Mansour Journal, 1 - 29.

Vill´an, R., Voloshynovskiy, S., Koval, O., Vila, J., Topak, E., Deguillaume, F., et al. (2006). Text Data-Hiding for Digital and Printed Documents: Theoretical and Practical Considerations. Stochastic Information Processing (SIP), 15-26.