Research Journal of Applied Sciences 11 (10): 1054-1059, 2016
ISSN: 1815-932X

# Text Steganography in Excel Documents Using Color and Type of Fonts

Samraa A. Al-Asadi and Wesam Bhaya
University of Babylon, College of Information Technology, Babil, Iraq

**Abstract:** Steganography is the science of hiding data within a cover object in order to keep the secret message invisible without affecting the integrity of the cover object, so the other individuals fail to recognize the presence of the secret message. This study concerns with the steganography in the MS Excel documents. We propose a new steganography method to hide data in the Excel sheets by changing both the font color and font type of each character of the text within each cell. The proposed method will hide two information bits -that belong to two successive characters within the secret message- in one character within a Excel cell by using both font color and font type. The first bit will be conceal by using two different values for the same color (for example: for Red color, two values can be used, 255 to conceal the bit "1" and the value 254 to conceal the bit "0") and the second bit will be conceal by using the font types that have two different values but the same appearance (for example: the font type "Arial" is used to represent the information bit "1" and the font type "Microsoft Sans Serif" is used to represent the information bit "0"). Using these two font attributes make all the text with the color "red" and with the same appearance but the value of the red color in each cell is really different and so the font type. Also the proposed approach distributes all the bits of secret message randomly all over the Excel sheet depending on a geometric manner making it difficult to extract the secret message back.

**Key words:** Information hiding, steganography, excel document, color and type of fonts, Iraq

## INTRODUCTION

Steganography is a technique of hiding information within a cover media so that no one can recognize their existence. Steganography has been originated since the time of the ancient Greeks and the word "steganography" means "concealed writing" from the Greek words "steganos" meaning "covered or protected" and "graphein" meaning "to write" (Saber and Awadh, 2013). There are many popular stenographic media like text, image, audio and video and Text is the most difficult media because the structure of the text file is exactly the same to what we observe and there is no redundant information can be used to hide data, in contrast to pictures or sounds cover media (Shahreza and Shahreza, 2010).

Steganography has the objective of keeping the secret message invisible without affecting the cover object (Bhaya, 2011), so it adds a new feature to the secrecy which called invisibility. This make the steganography differ from cryptography which only ensures the confidentiality of the message content (Bhaya *et al*., 2013). Steganography has two branches, the first one is the "Protection against Detection" and the other is the "Protection against Removal". Protection against Removal means placing a hidden trademark in a picture, music or in software and it is the technique called Watermarking (Rabah, 2004). Figure 1 shows a general view of steganography system (Bhaya, 2011).

Steganography method in the embedding part has two inputs (the cover and the secret message) and the output is the Stego medium which looks like the same as the cover but with the invisible secret message in it, where the key is the applied method of the steganography (Gutub and Fattani, 2007).

**Literature review:** The related published works in steganography for the text document are:

- Khairullah (2014) proposed a method of steganography in a financial report since it has a huge number of numeric data. The idea behind this approach is to pad additional zeros before the number and padding additional zeros after or before the fractional part of the number. These added zeros used to conceal the secret bits
- Esraa proposed a steganography method for Arabic text, by using two Diacritics ('Fathah' and 'Kasrah') to conceal information bits (0,1) by keeping or removing these diacritics from the cover text (Ahmadoh and Gutub, 2015)

**Corresponding Author:** Samraa A. Al-Asadi, University of Babylon, College of Information Technology, Babil, Iraq

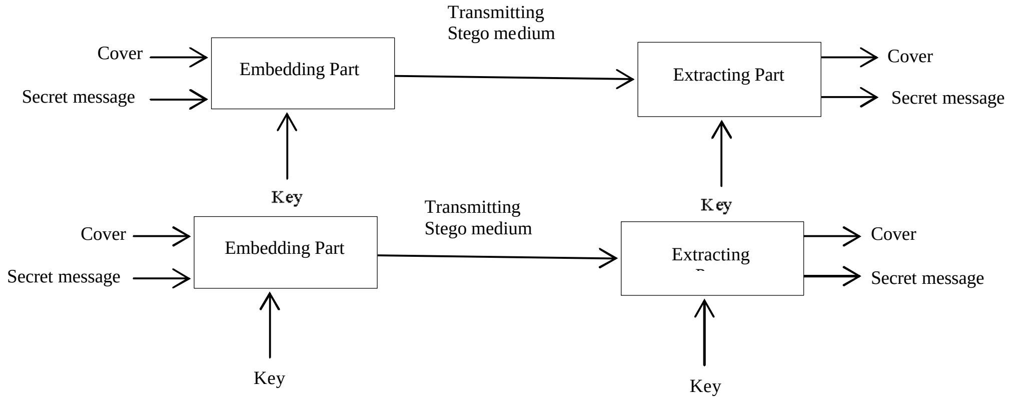


Fig. 1: The Steganography system

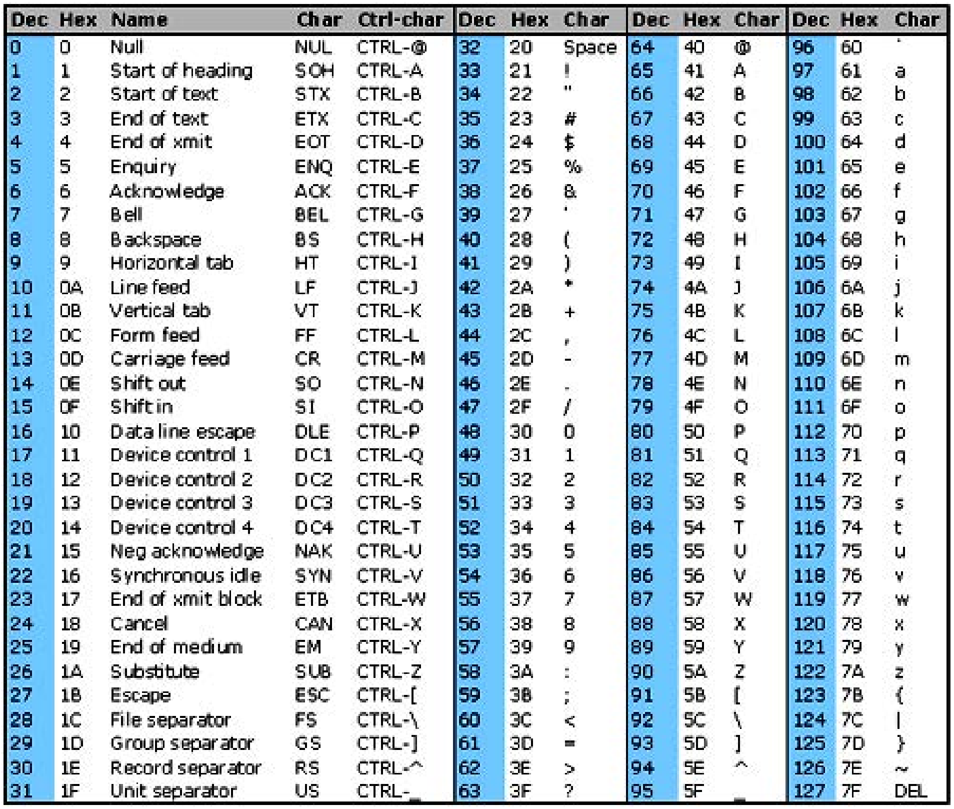
- Bhaya (2011) propose a method depending on using similar Font types of English letters in hiding message by changing the font to another similar font. The first method is about concealing the secret message in a mobile phone SMS (Simple Message Service (SMS) where the mobile phones uses two default font's types which they are "Proportional" and "System", every secret letter will be concealed in five letters from the cover text since each secret letter is coded into five bits
- Bhaya *et al.* (2013) also propose a method depending on using similar Font types, where the secret message will be concealed in only the capital letters of the cover document, by first determining the font of the cover document in order to get its similar fonts. Then each letter of the secret message will be coded into three codes (0,0,0), (0,0,1)… to (3,3,3), so this approach will hide only the each secret letter within three capital letters form the cover text (Bhaya *et al.*, 2013)
- Wael presents a steganography system for Arabic text. They get benefit of having points within more than half the letters of the Arabic text. Their scheme conceals the bit "1" within the pointed letters, while the bit "0" is concealed using the un-pointed letters. In addition, their steganography scheme uses extension characters beside the letters to note the specific letters holding the hidden bits (Gutub and Fattani, 2007)
- Khairullah (2009) propose a steganography method in MS Word documents by controlling the foreground color of the indivisible characters like the space or the carriage return, these characters within the documents are not reflected or viewed
- Suhad and Abdulraheen (2015) proposed a steganography method for Arabic text by using the grammar of Arabic language as a logical term, B+ tree is used for storing and Augmented Transition Network (ATN) used for parsing (Suhad and Abdulraheem, 2015)
- Khan *et al.* (2015) proposed a method for Czech text, this method using the un-pointed letters with extension to conceal the bit "0", while the bit "1" is concealed within the pointed letters that have an extension (Tiwari and Sahoo, 2011)

While some related published works in steganography for the Excel documents are:

- Yang *et al.* (2011) proposed a text rotation technique. This method hides the information by slightly rotating the angle of the cell to embedded information. By measuring the angle of each cell, the secret information will be retrieved
- Salman and Akeel propose a steganography technique for hiding data using Unicode system characteristics method. In Arabic and Persian Language, there are two different codes for seven numbers {9, 8, 7, 3, 2, 1 and 0} in the Unicode table. The seven numbers of {9, 8, 7, 3, 2, 1 and 0} have the same shape but different codes in Unicode standard (Saber and Awadh, 2013)
- Tiwari and Sahoo (2011) propose a three techniques methodology for embedding the secret data with three layers of security and robustness

## MATERIALS AND METHODS

The proposed method support concealing the (alphabet A-Z, a-z, 0-9 and all other symbols), each of



| Dec | Hex | Name | Char | Ctrl-char | Dec | Hex | Char | Dec | Hex | Char | Dec | Hex | Char |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 0 | Null | NUL | CTRL-@ | 32 | 20 | Space | 64 | 40 | @ | 96 | 60 | ` |
| 1 | 1 | Start of heading | SOH | CTRL-A | 33 | 21 | ! | 65 | 41 | A | 97 | 61 | a |
| 2 | 2 | Start of text | STX | CTRL-B | 34 | 22 | " | 66 | 42 | B | 98 | 62 | b |
| 3 | 3 | End of text | ETX | CTRL-C | 35 | 23 | # | 67 | 43 | C | 99 | 63 | c |
| 4 | 4 | End of xmit | EOT | CTRL-D | 36 | 24 | $ | 68 | 44 | D | 100 | 64 | d |
| 5 | 5 | Enquiry | ENQ | CTRL-E | 37 | 25 | % | 69 | 45 | E | 101 | 65 | e |
| 6 | 6 | Acknowledge | ACK | CTRL-F | 38 | 26 | & | 70 | 46 | F | 102 | 66 | f |
| 7 | 7 | Bell | BEL | CTRL-G | 39 | 27 | ' | 71 | 47 | G | 103 | 67 | g |
| 8 | 8 | Backspace | BS | CTRL-H | 40 | 28 | ( | 72 | 48 | H | 104 | 68 | h |
| 9 | 9 | Horizontal tab | HT | CTRL-I | 41 | 29 | ) | 73 | 49 | I | 105 | 69 | i |
| 10 | 0A | Line feed | LF | CTRL-J | 42 | 2A | * | 74 | 4A | J | 106 | 6A | j |
| 11 | 0B | Vertical tab | VT | CTRL-K | 43 | 2B | + | 75 | 4B | K | 107 | 6B | k |
| 12 | 0C | Form feed | FF | CTRL-L | 44 | 2C | , | 76 | 4C | L | 108 | 6C | l |
| 13 | 0D | Carriage feed | CR | CTRL-M | 45 | 2D | - | 77 | 4D | M | 109 | 6D | m |
| 14 | 0E | Shift out | SO | CTRL-N | 46 | 2E | . | 78 | 4E | N | 110 | 6E | n |
| 15 | 0F | Shift in | SI | CTRL-O | 47 | 2F | / | 79 | 4F | O | 111 | 6F | o |
| 16 | 10 | Data line escape | DLE | CTRL-P | 48 | 30 | 0 | 80 | 50 | P | 112 | 70 | p |
| 17 | 11 | Device control 1 | DC1 | CTRL-Q | 49 | 31 | 1 | 81 | 51 | Q | 113 | 71 | q |
| 18 | 12 | Device control 2 | DC2 | CTRL-R | 50 | 32 | 2 | 82 | 52 | R | 114 | 72 | r |
| 19 | 13 | Device control 3 | DC3 | CTRL-S | 51 | 33 | 3 | 83 | 53 | S | 115 | 73 | s |
| 20 | 14 | Device control 4 | DC4 | CTRL-T | 52 | 34 | 4 | 84 | 54 | T | 116 | 74 | t |
| 21 | 15 | Neg acknowledge | NAK | CTRL-U | 53 | 35 | 5 | 85 | 55 | U | 117 | 75 | u |
| 22 | 16 | Synchronous idle | SYN | CTRL-V | 54 | 36 | 6 | 86 | 56 | V | 118 | 76 | v |
| 23 | 17 | End of xmit block | ETB | CTRL-W | 55 | 37 | 7 | 87 | 57 | W | 119 | 77 | w |
| 24 | 18 | Cancel | CAN | CTRL-X | 56 | 38 | 8 | 88 | 58 | X | 120 | 78 | x |
| 25 | 19 | End of medium | EM | CTRL-Y | 57 | 39 | 9 | 89 | 59 | Y | 121 | 79 | y |
| 26 | 1A | Substitute | SUB | CTRL-Z | 58 | 3A | : | 90 | 5A | Z | 122 | 7A | z |
| 27 | 1B | Escape | ESC | CTRL-[ | 59 | 3B | ; | 91 | 5B | [ | 123 | 7B | { |
| 28 | 1C | File separator | FS | CTRL-\ | 60 | 3C | < | 92 | 5C | \ | 124 | 7C | \| |
| 29 | 1D | Group separator | GS | CTRL-] | 61 | 3D | = | 93 | 5D | ] | 125 | 7D | } |
| 30 | 1E | Record separator | RS | CTRL-^ | 62 | 3E | > | 94 | 5E | ^ | 126 | 7E | ~ |
| 31 | 1F | Unit separator | US | CTRL-_ | 63 | 3F | ? | 95 | 5F | _ | 127 | 7F | DEL |

Fig. 2: Ascii codes table

these characters will be coded into 7 bits Ascii code starting from "0000000" and ending with "0111111" while the code "1111111" that denotes the "Del" character will always set to the end of the code stream to denote the end of the secret message -this special code has the benefit at the extracting process- while there is no special code for the beginning of the code stream of the secret message since it will be set at the beginning of the cover document. The "Del" character cannot be used within the secret message. The Ascii codes are shown in the Fig. 2.

Using the Excel sheet has the benefit of distributing the secret bits through the cells, in contrast to the MS Word or PPT documents where the embedding of bits is somehow done sequentially. In this approach; each character within each cell can hide 2 bits from the secret bit stream (these two bits belong to two different characters) by controlling both the font color and font type. The manner for hiding the bits is shown in Fig. 3, where the cells are indexed according to its turn in the steganography process (both embedding and extracting processes) and for each cell we use all the characters within it where each character conceals two bits. (This represents the key of the algorithm).

| | | | | | | |
|---|---|---|---|---|---|---|
|  | 1 | 3 | 5 | 7 | 9 | 11 |
| 2 |  | 13 | 15 | 17 | 19 | 21 |
| 4 | 14 |  | 23 | 25 | 27 | 29 |
| 6 | 16 | 24 |  | 31 | 33 | 35 |
| 8 | 18 | 26 | 32 |  | 37 | 39 |
| 10 | 20 | 28 | 34 | 38 |  | 41 |
| 12 | 22 | 30 | 36 | 40 | 42 |  |

Fig. 3: Sample of excel sheet shows how to reach each cell for using its characters in the steganography process

The proposed method will hide two information bits either "00", "01", "10" or "11" in each character within each cell by using two different values for the same color, for example: for Black color, two values can be used, 0 to represent the information bit "1" and the value 1 to represent the information bit "0". This is for the first bit. The second bit will be embedding using two different font types that have the same shape for example the font type "Arial" is used to represent the information bit "1" and the font type "Microsoft Sans Serif" is used to represent the information bit "0").

Using these values will make all the text with the color" Black" and with the same appearance but the value of the Black color and the font type for each character in each cell is really different. This makes the proposed method hard to be detected by human vision.

The proposed steganography method has been implemented using Visual Basic programming language and by using the MS Excel documents. The proposed method has two parts:

**Hiding (Embedding) Part:** As shown in Fig. 4, this part has the following steps:

- Step 1: Create an Excel document and fill it with the plain text or the cover message, or found an excel document that is previously created and filled with the plain text
- Step 2: Coded the secret message (each character in 7 bits Ascii code) and tailing this stream of bits with the special code"1111111"
- Step 3: for each two bit indexed (i) and (i+7) -each two bits (i) and (i+7) of a successive characters within the secret message will be used at the same time- in the code stream as shown in Fig. 5 this step will be repeated:
  - Find a character in a cell according to he cell's turn as shown in Fig. 3
  - Embedding the first bit by controlling the color and the second bit by controlling the type of the font of the same character
  - Save the final excel document (that containing the plain text and the invisible secret message)

Extracting (Retrieval) Part: as shown in Fig. 6, this part has the following steps:

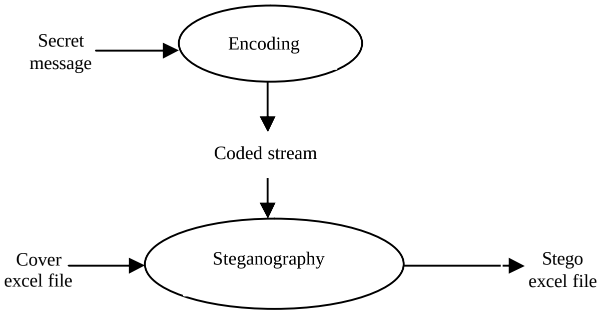


Fig. 4: The embedding part of the proposed steganography method

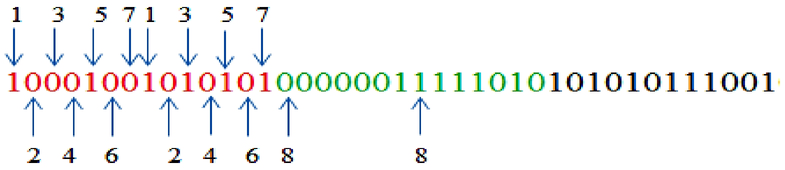


Fig. 5: Secret bits order in the processing

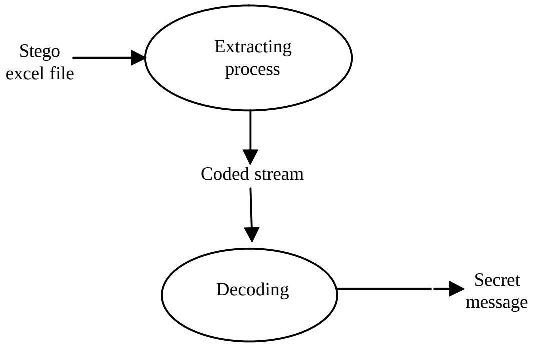


Fig. 6: The Extracting part of the proposed steganography method

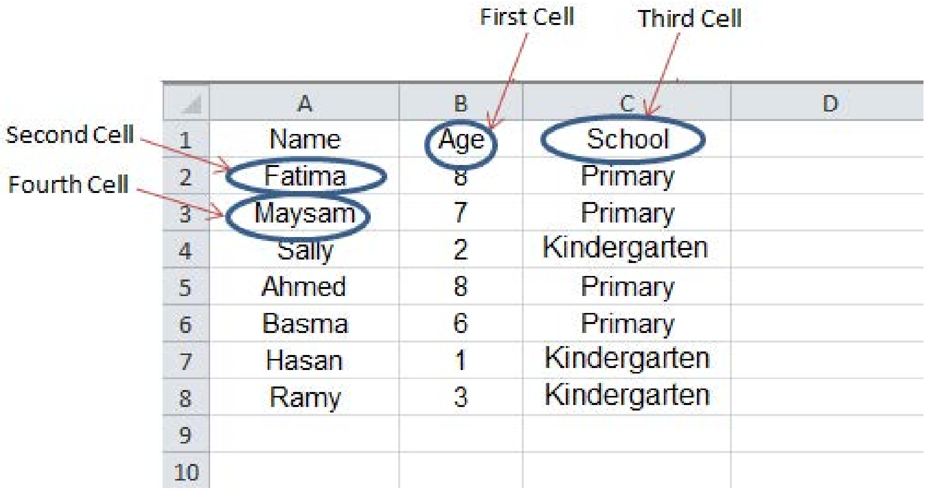
- Step 1: Open an excel document which have the embedded but invisible secret message
- Step 2: According to the manner for hiding the bits that shown in Fig. 3, this approach will reach each cell according to its turn and for each character within it, a two bit will be retrieved-the first one is depending on the character font color and the second one depending on the character font type-each bit will be included in the message coded stream according to its real position as shown in Fig. 5 (Step 2 will be repeated until the special code “1111111” will be found)
- Step 3: Retrieve the secret message from the coded stream

## RESULTS AND DISCUSSION

**Case study:** Let us propose that the secret message is “Hello” and we have an Excel sheet with the color Black = 0 and Font type=”Arial”. When embedding the secret message in that Excel document, it will be coded first into this coded stream -according to the Ascii code table shown previously in Fig. 1- “100100011001011101100110 11001101111**1111111**". Figure 7 shows the turn of each cell whereas. Figure 8 shows controlling of the characters’ font type and color depending on the bit (0 or 1), where the black color value=1 is used to represent the bit (1), while the black color value =0 is used to represent the bit (0) and the font type “Arial” is used to represent the bit (1), while the font type “Microsoft Sans Serif” is used to represent the bit (0).

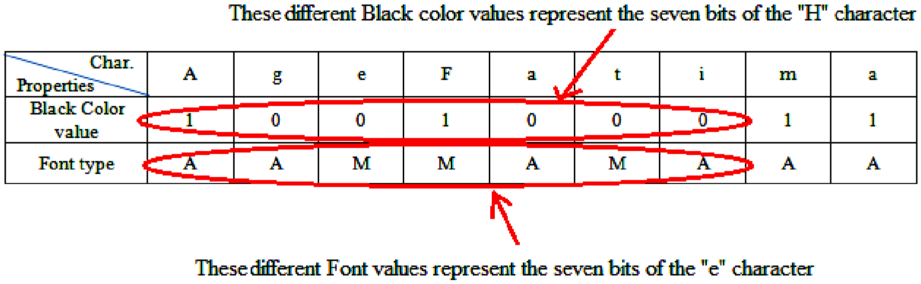
At the retrieval process, form each character within each cell as its turn in the processing, two bits will be extracted depending on font color and font type. For the previous example and for the first two cells, this stream will be retrieved

“110100100100011111”. Figure 9 shows the retrieval of secret message from this stream of bits.



|  | A | B | C | D |
|---|---|---|---|---|
| 1 | Name | Age | School |  |
| 2 | Fatima | 8 | Primary |  |
| 3 | Maysam | 7 | Primary |  |
| 4 | Sally | 2 | Kindergarten |  |
| 5 | Ahmed | 8 | Primary |  |
| 6 | Basma | 6 | Primary |  |
| 7 | Hasan | 1 | Kindergarten |  |
| 8 | Ramy | 3 | Kindergarten |  |
| 9 |  |  |  |  |
| 10 |  |  |  |  |

Fig. 7: The stego excel document with the embedded and invisible secret message


These different Black color values represent the seven bits of the "H" character

| Char. / Properties | A | g | e | F | a | t | i | m | a |
|---|---|---|---|---|---|---|---|---|---|
| Black Color value | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 1 |
| Font type | A | A | M | M | A | M | A | A | A |

These different Font values represent the seven bits of the "e" character

Fig. 8: Font type and color for the characters in both first and second used cell in the steganography process

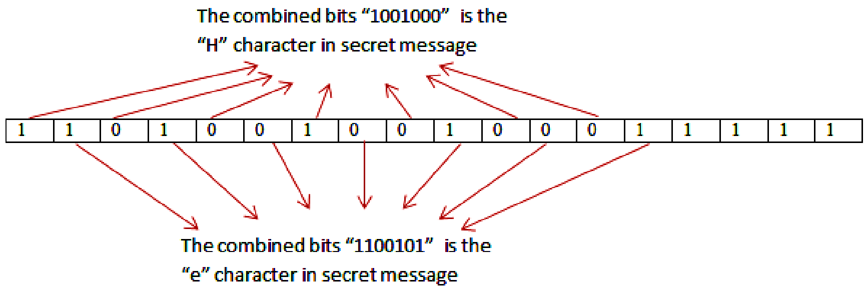

The combined bits "1001000" is the "H" character in secret message
1 1 0 1 0 0 1 0 0 1 0 0 0 1 1 1 1 1
The combined bits "1100101" is the "e" character in secret message

Fig. 9: Retrieval process of characters from the extracted bit stream

## CONCLUSION

This research develops a new approach for embedding a text in an excel document, it depends on using two similar colors with different values for the two bits (0, 1) and using two similar font type with different font's type name for the two bits (0,1). The approach can hide every two secret characters within 7 characters of the cover text. This similarity in colors and font makes this approach difficult to be recognized by human vision and as a result, the embedded secret message will not be detected easily.

Using this approach, it will be difficult to predict the original secret text since the secret bits are distributed through the cells of the Excel document in a geometric non-sequential manner.

The secret message bits are also processed in non-sequential manner by picking up two bits that belong to two successive characters and hide them in the same character in the Excel sheet.

This approach like any other approaches is vulnerable to the Man-in-the-Middle attack, since any steganography method is sensitive to any change in the document properties or its content, so if the attacker make any change, the secret message will not be retrieved correctly. This is a drawback to any steganography method.

## REFERENCES

Ahmadoh, E.M. and A.A.A. Gutub, 2015. Utilization of two diacritics for Arabic text steganography to enhance performance. Lecture Notes Inf. Theory, 3: 42-47.

Bhaya, W., A.M. Rahma and A.N. Dhamyaa, 2013. Text steganography based on font type in ms-word documents. J. Comput. Sci., 9: 898-904.

Bhaya, W.S., 2011. Text hiding in mobile phone simple message service using fonts. J. Comput. Sci., 7: 1626-1628.

Gutub, A.A.A. and M.M. Fattani, 2007. A novel Arabic text steganography method using letter points and extensions. World Acad. Sci. Eng. Technol., 21: 28-31.

Khairullah, M., 2014. A novel text steganography system in financial statements. Int. J. Database Theory Appl., 7: 123-132.

Khairullah, M.D., 2009. A novel text steganography system using font color of the invisible characters in Microsoft Word documents. Proceedings of the Second International Conference on Computer and Electrical Engineering (ICCEE'09), December 28-30, 2009, IEEE, Sylhet, Bangladesh, ISBN: 978-1-4244-5365-8, pp: 482-484.

Khan, S., R. Sankineni, P. Balagurunathan, N.S.D. Shree and A. Balasubramanian, 2015. Czech text steganography method by selective hiding technique. Proceedings of the World Congress on Engineering, July 1-3, 2015, WCE, London, England, ISBN:978-988-19253-4-3, pp: 1-4.

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

Saber, A.S. and W.A. Awadh, 2013. Steganography in MS excel document using unicode system characteristics. J. Basrah Res. Sci., 39: 10-19.

Shahreza, M.H.S. and S.M. Shahreza, 2010. Arabic-Persian text steganography utilizing similar letters with different codes. Arabian J. Sci. Eng., 35: 213-222.

Suhad, M.K. and A.A. Abdulraheem, 2015. Using a parser for steganography purpose. Eng. Tech. J., 33: 654-667.

Tiwari, R.K. and G. Sahoo, 2011. Microsoft Excel file: A steganographic carrier file. Int. J. Digital Crime Forensics (IJDCF.), 3: 37-52.

Yang, B., X. Sun, L. Xiang, Z. Ruan and R. Wu, 2011. Steganography in Ms Excel document using text-rotation technique. Inform. Technol. J., 10: 889-893.