EasyChair Preprint

№ 148

Text Steganography in Font color of MS Excel Sheet

Husam Ibrahiem Alsaadi, Maad Kamal Al-Anni and
Rafah M. Almuttairi

May 21, 2018

Text Steganography in Font color of MS Excel Sheet

Husam Ibrahiem Alsaadi, PhD student, Altinbas University, Istanbul, Turkey. (email: husam.alssadi@ogr.altinbas.edu.tr)

Dr. Maad Kamal Al-Anni, Software and Networking Engineering College, Al-Iraqia University, Ministry of Higher Education and Scientific Research, Iraq, (email:- maadk_anni@live.com).

Assistant Prof. Dr. Rafah M. Almuttairi, Department of information networks, College of Information Technology, University of Babylon, Ministry of Higher Education and Scientific Research, Iraq,(email:, rafahmohammed@gmail.com,.

Abstract

One of the ways to maintain the security of data transfer over the network is to encrypt the data before sending it, for the purpose of increasing data privacy and its importance, **Steganography**, is used to hide encrypted data in various documents, such as sheets of MS Excel sheets.

The purpose of the encrypted data packaging during transport from one place to another is, to convince an unauthorized person, nothing has not been added to the transport data. The art of concealment is embodied hidden words to be passed confidentially within a file without influence on the properties of the file and this can't be anyone's hand is to know that there is confidential data encapsulated by transport data. Whereas, the Microsoft Excel file size being used to hide data remains constant in spite of the file containing the confidential data.

The secret for the survival of file size unchanged despite the addition of encrypted data, is that confidential data is added to the value of the color of the font used in the cells, leading to the font color change slightly.

Previous research used the Excel file to hide data in different ways, but the data masking and packaging by changing the color content writing is what distinguishes our research and make it alone in this area.

Keywords: Steganography; Excel documents; Font color.

Introduction:

There are a lot of ways and means that are used by some people to sabotage the property of others, whether tangible or softcopy such as data stored in computer devices. In modern era or the so-called information era, the attempts to subvert the data have grown more than ones trying to destroy sensory value of property, this is what has enriched research on ways in which they can hide information about who does not possess them and put it in the safe. As well as the means by which to detect any hidden information illegally.

When talking about Information hiding, the first thing that comes to mind is information encryption process (Cryptography) or converted to another format is understood only by the code

or the key (key). But this does not mean that encryption is the only way in which they can hide information from third parties. In fact, the meaning of Steganography is only one way to hide the information, which is often used by information thieves, to steal sensitive information.

Steganography is English term, means to hide information by packaging it in different ways. The used methods might be illegal, used for the purpose of convey the secret data from one place to another place, but without knowing anybody to do so.

Encryption (Cryptography) is also used to hide the content of the message (but everyone knows its existence), while Steganography is used to hide the existence of the message already. For example, a person, may use the electronic picture to transfer secret text messages (or even hidden photos) to someone else without knowing anyone.

All who look from outside, think that the two people exchanging normal images, while these images loaded with invisible hidden messages. But how can an image that contains these texts or messages? The answer lies in, the electronic image is only a set of binary points "pixels" , which based on the pattern it may contain a number of binary data bits which is unimportant or unused. So the person who wants to hide the secret message in a picture, unimportant bits should be exploited and filled by his message. it should be noted that Steganography is quite different from the watermark or the (water mark), in Steganography be concealment of illegally for the purpose of illegal, while in watermark mostly used to save copyright.

Choosing a transmission medium to hide a secret messages, the carrier of the secret message, is the most important thing to success transfer the confidential data to other party with totally invisible way . The carrier of secret message which contains a lot of unused data (bits unused or redundant) is ideal for transport.

Usually determine the carrier's ability to transfer the hidden information through, proportion of information that can be hidden to the clear phenomenon of information. For example, if there is a carrier through which only 1 byte can be added as hidden data for each 100 bytes of clear data, the conductivity of this carrier is 1%.

In addition to the conductivity of the carrier, there is another important feature, called the degree of disappearances. If the information concealed in the carrier is very difficult to be discovered, the carrier is better.

The most important qualities of a good carrier is, the secret message which is hidden won't change during transport. Examples of media that can be used as a carrier of the secret message are, images (as in the previous example), audio files and video, data transmitted across networks by different protocol (TCP / IP, UDP, ICMP ... etc).

There are many programs that have been developed to hide messages in different media, such as JSteg program that hides messages in image JPEG files, and Hidden Secrets program that hides messages in different files, such as JPEG files, BMP, HTML, PNG type, WAV.

Related Works

Steganography is the ancient art and young science of hidden communication. A broad definition of the subject includes all end eavours to communicate in such a way that the existence of the message cannot be detected. Steganography is a significant means that secret information is embedded into cover data imperceptibly for transmission, so that information cannot be easily aware by others. Text Steganography is low in redundancy and related to natural language rules these lead to limit manipulation of text, so they are both great challenges to conceal message in text properly and to detect such concealment. The secret message was encoded and embedded as similar fonts in capital Letters of cover document. Proposed text steganography method can works in different cover documents of different font types[1].

Earlier information hiding methods merely embed payload (external information) into a cover (e.g., text document, image and audio) and in recent years, specialized data hiding methods are proposed to serve specific purposes. For instance, in steganography, the cover content is carefully manipulated to encode payload while aiming to conceal the very existence of the encoded information [2].Steganography literally means “covered writing” and is the art of hiding the very existence of a message. A message is the information to be hidden, anything that can be embedded into a bit stream. Together the cover carrier and the embedded message create a steganocarrier. Hiding information may require a stegano key which is additional secret information, such as a password, required for embedding the information [3].

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 "grapheid' meaning "to write" [4].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 obscene and there is no redundant information can be used to hide data, in contrast to pictures or sounds cover media [5].

A propose new approach for information hiding using inter-word spacing and inter-paragraph spacing as a hybrid method in [6].

Anew approach for steganography in Microsoft Word documents, the main idea is that setting any foreground color for invisible characters such as the space or the carriage return is not reflected or viewed in the document[7].

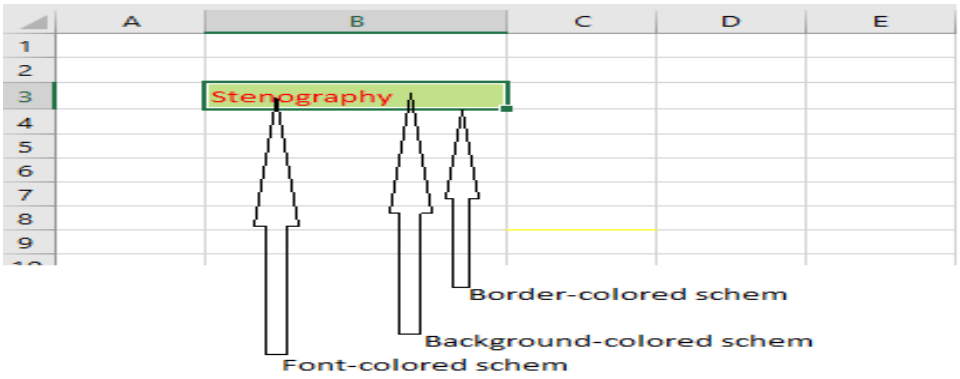
Steganography is the science of hiding data within a cover object in order to keep the secret message invisible without affecting the integrity of the cover object, so the other individuals fail to recognize the presence of the secret message. Concerns study with the steganography in the MS Excel documents 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. Also 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[8].

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[9].
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[10].
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 [11]. 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 information by the eavesdropper. Robustness relates to the amount of modification the stego-medium can withstand before the adversary destroys the hidden information [12]. It should be seeking for the appropriate balance between the three aspects according to the specific requirement.

Materials and methods

By default, the Microsoft office has its own color schema of the cell/cells in MS Excel as Red Green Blue RGB for three Color Minutia, 1st is Font formatting, 2nd is Border Formatting, and last one is foreground Formatting, as it is shown in fig.1, this schema allows us to utilize this



technique in order to apply the text stenography using the less significant bit (LSB).

Fig. 1: MS-Excel cell's color schema

The LSB is the lowest significant bit in the byte value of the three Colored Minutia(RGB).The LSB based RGB steganography embeds the secret message in the least significant bits of RGB values of the Cover Excel File (CEF) as illustrated by Fig. 2. Whereas the intended text for steno graphing is called as a secret text that would be embed by either of three colored Minutia, in this

article we are going to select the font formatting in order to hide the text secret message, due to the length of secret message is rather enough to be apply for only this formatting schema, besides it could be applicable for the rest of other formatting.

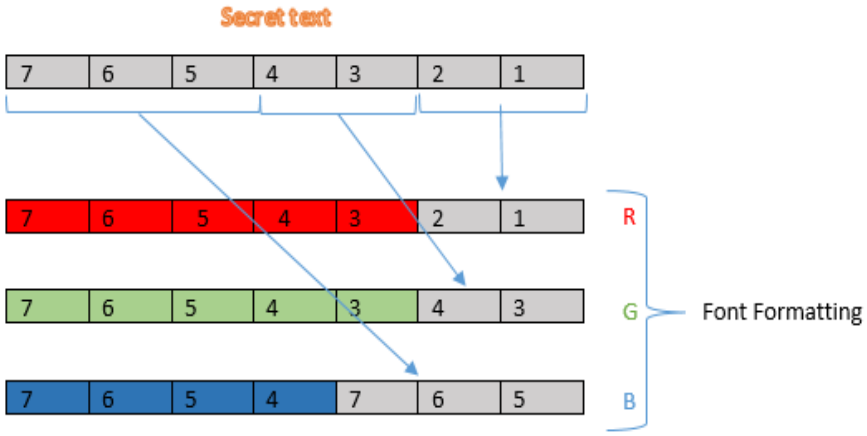


Fig. 2: LSB Algorithm

The secret text is presenting in the ASCII Code presentation (alphabet a-z or A-Z, number 0-9, all other symbols) while later on it is coding in bit format to prepare them to be scattered to excel cells in order to achieve the text stenography into cell color schema, Fig. 3 shown the decimal or hexadecimal numbers against the character's symbols in Fig. 3.

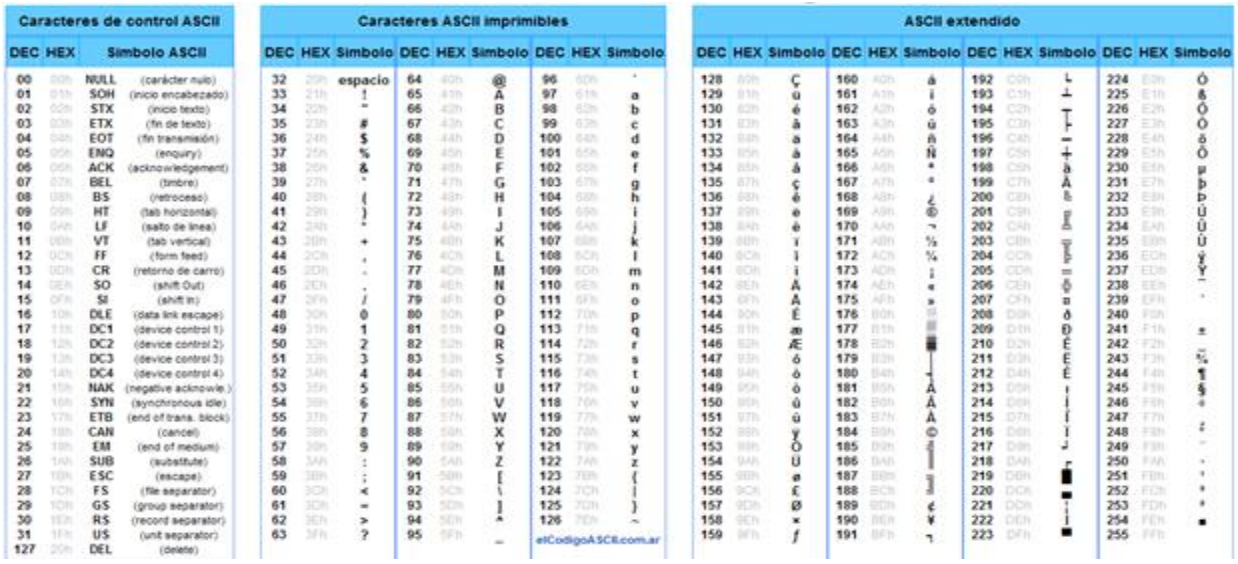


Fig. 3: ASCII Code Table

Some of ASCII Code characters is the actual coding of secret message characters that will be coded into 8 bits ASCII code could be starting from "0000000" equivalent to 0 Decimal no ($0*2^7+0*2^6+0*2^5+0*2^4+0*2^3+0*2^2+0*2^1+0*2^0$) up to "1111111" equivalent to 127 Decimal no($1*2^7+1*2^6+1*2^5+1*2^4+1*2^3+1*2^2+1*2^1+1*2^0$), due to this fact, the code "1111111"

equivalent to “DEL” character in the ASCII Code table hence it is not necessary to reach at the end of code stream table to end up with the extracting process, in this article, instead of doing the longer process of code stream that will be unbeneficial in the processing time point of view, then we are going to adequate by only presenting 7 bits(but it could be 8 bits or 6 bits is also applicable) for binary coding and having the rang of transmitting characters that is indicated in the header information by 9999 decimal (also could be 99999 or more depending on the secret message length) in which it be able to identify the ending of secret message while doing retrieval process, we are going to deal with the end of secret massage through 2^{nd} up to 5^{th} reserved cells (denoted by the number of transmitting characters)as well as rest of cells are the actual cells available for secret message shown in Fig. 4, hence the first one is allocated for encoding process into either 8, 6 , or 7 bits, the 2^{nd}one till 5^{th} one is resolved for the number of transmitting characters for secret message, the 6^{th} one for color minutia, it could be either, 1 denoted to the 1^{st} is Font formatting, 2 denoted to the 2^{nd} is Border Formatting, or 3 denoted to the 3^{rd} is foreground Formatting, in our schema is 1 as a Font Formatting, while the beginning of coding stream of the secret message since it will be start immediately after the 6^{th}consecutive filled cell which be the number 7^{th} serially till the total characters number(n) + 6(header information), suppose we are going to send the secret message of the length of 175 chars, then the number of cells required to do the stenography process for such an example is N+6 or 175+6 which is equal to 181 cells(actual filled cells), Fig. 5 shown the header information as well as available locations for secret message.

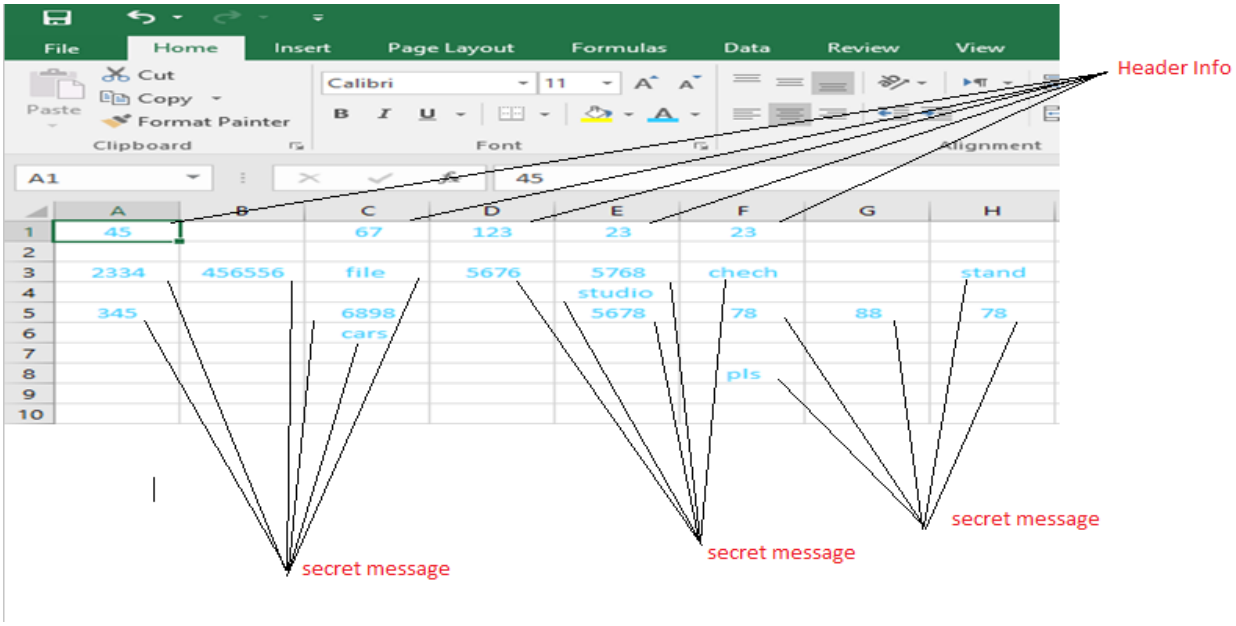


Fig. 4 resolved cells and cells available for secret message

1st cell	2nd cell	3rd	4th	5th	6th	7th	.....	Nth
6 bits (2 R, 2 G, 2 B) 7 bits (2 R, 2 G, 3 B) 8 bits (2 R, 3 G, 3 B)	0-9 Secret message length	0-9 Secret message length	0-9 Secret message length	0-9 Secret message length	1 – 2 – 3 Color minutia	Available for secret message	.....	Nth
Header information								

Fig. 5 Header Information and secret message

The length of are stored in 2nd cell till 6th cell while it is going to retrieved during the decoding using the Mode operation as shown in the pseudo-code algorithm-1, for example, if we have the secret message of the length 4 bits and the number of stored characters is 0175, then '0' stored in 2nd cell, '1' in 3rd cell, '7' in 4th cell, and the next one which is '5' stored ultimately in 5th cell, likewise it is capable to be embedded the max number is 9999 decimal such as in our example, hence the last one is the 6th cell that is reserved for color minutia(in our example would be '1' for the whole approach as the font formatting).

Psedo-Code Algorithm-1-(Mode Operation for 4 bits):-

```
1-  X=no;          " the retrieved number from the excel cells"
2-  Di=0;                  " dividend number "
3-  Mo=0;              " mode results of the 4th digits "
4-  For k = 1 : 4 do       " the loop of 4 steps '9999'"
5-  Di = fix (x/10);         "division operation"
6-  Mo=mode (x, 10);       " mode operation "
7-  Bi=dec2bin(mo, 4);     "converting decimal to binary "
8-  A(k, :)= (mo);
9-  X=di;
10- End for                  "end for loop "
```

Before starting to scatter the secret message into Cover Excel File(CEF), the Auto-Algorithm for transferring the Secret Message to another form by using the transposition matrix and flipping box in order to increase the diffusion and impossibility of guessing among the hiding message over CEF, below Fig. 6 a and Fig. 6 b (including five steps before Scattering operation)shows the pre-operations that precede the scattering operation into filled cells into CEF, other pre-process of the auto-algorithm is being verified whether the CEF suitable for Hiding process or not by calculating the total number of filled cells comparing it with the actual secret message to be secreted into CEF(total filled cells >= no of characters of secret message).

A1	A2	A3	A4	A5	A6	A7
B1	B2	B3	B4	B5	B6	B7
C1	C2	C3	C4	C5	C6	C7
D1	D2	D3	D4	D5	D6	D7

1st step

2nd step

A1	B1	C1	D1	A2	B2	C2	D2	A3	B3	C3	D3	A4	B4	C4	D4	A5	B5	C5	D5	A6	B6	D6	D6	A7	B7	C7	D7

3rd step

B6	C6	D6	A7	B7	C7	D7
C4	D4	A5	B5	C5	D5	A6
D2	A3	B3	C3	D3	A4	B4
A1	B1	C1	D1	A2	B2	C2

Fig. 6 a (columnar and concatenation transfer)

4th step5th step

B6	C6	D6	A7	B7	C7	D7
C4	D4	A5	B5	C5	D5	A6
D2	A3	B3	C3	D3	A4	B4
A1	B1	C1	D1	A2	B2	C2

Flipping Operation

D7	C7	B7	A7	D6	C6	B6
A6	D5	C5	B5	A5	D4	C4
B4	A4	D3	C3	B3	A3	D2
C2	B2	A2	D1	C1	B1	A1

Fig. 6 b (flipping operation of the 7 bits)

After getting the result of the columnar and concatenation transfer, the next step is to do the flipping operation from the right most side to the left most side of the throughput transfer by 7 bits of each step as shown in Fig.6 b, whereas the first row of the 1st step done by thecolumnar and concatenation transfer that is deferred from the first row of the 5th step done by flipping operation as shows in fig. 7, now each 7 bits is ready to scathe into Cover Excel File(CEF), we do the segmentation of the latter result as $2^6 2^5$ embedded into the LSB of Red Minutia, $2^4 2^3$ embedded into the LSB of Green Minutia, and $2^2 2^1 2^0$embedded into the LSB of Blue Minutia, the process of scattering till there is no characters left over or if the loop reach to the end of loop (N+header length, N+6), notify that the process of decoding is done by the same way but in the reversed steps.

The main advantage for scathing the secret message upon the CEF is making the brute force more confusing and has no possible to visualize the hiding message in normal circumstance, it is somehow rather than embedding it into either MS Word or PPT while the latter done sequentially, in this approach: each cell can hide a character with coding process (6 bits, 7 bits, and 8 bits) as well as it has three manner of hiding, 1st within font color, 2nd within border color, and 3rd within foreground color, each of them capable to embedded its secret message of a character, therefore this approach has advantage to distribute three characters within one cell,

suppose we have a secret message of 500 character, then it is mandatory to prepare a CEF for the number of cells at least 500/3 which is 167 cells proximately.

Noted that the first row as the result of the Flipping operation matrix is totally differed from that one in the first step of the columnar and concatenation operation as shows in Fig. 7, therefore the Matrix results of 5^{th} step during the flipping operation is scattered in CEF as row by row into each filled cell until the last filled cell, likewise this operation could be hidden each row of 7 bits into font color schema for every turn.

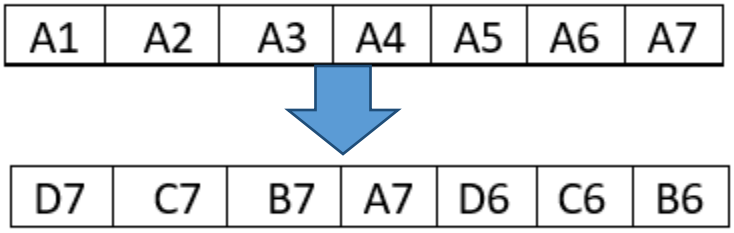


Fig. 7: first row as the results of flipping matrix

In contrary to the MS-Word and MS-PPT documents is done somehow sequentially, in MS-EXCEL has its properties to embedded the secret bits through CEF in such a way that is different arrangements to obfuscating the hidden message into CEF. Below are Fig. 8, the excel sheet and the pseudo-code algorithm -2 of searching filled cells into CEF.

					12/4/2003	
NAME		CAR NAME	MODEL	COLOR	WEEL	
Ahmed	34		2013	W	4	CLX
SAAD	50	mercedis			4	
REEM		honda	2003	G	4	
KAREEM	23	toyota	2017	G	4	
HUDA	34	honda	2009	B	4	

	A	B	C	D	E	F	G
1						1	
2	2		3	4	5	6	
3	7	8		9	10	11	12
4	13	14	15			16	
5	17		18	19	20	21	
6	22	23	24	25	26	27	
7	28	29	30	31	32	33	

Fig. 8: excel sheet shows the contains of cells and how to fetch each cell for using its color properties in the stenography

PSEDO-CODE ALGORITHM-2 FOR FETCHING THE NONE-EMPTY CELLS IN CEF FOR STENOGRAPHY PROCESS.

```
for R = firstFillRaw : lastFillRaw
for C = firstColumn: lastColumn
        myCell  = nn2an (R , C );
        Excel.Range(myCell).Select;
        myData= Excel.ActiveCell.Value;
if ~isnan(myData)
            fontColor= Excel.ActiveCell.Font.Color;
            cellColorBinary = dec2bin(fontColor, 24);
else
break
end
end
end
```

in this approach, each cell can hide 7 bits in the Font Color Formatting, two bits in the Red Color, two bits in the Green Color, last three in the Blue Color, likewise the manner for hiding the bits is shown in Fig. 8, where the cells are indexed according to its turn sequentially using the auto-program as shown in psedo-code Algorithm-2, the stenography process uses the cell`s turn both embedding and extracting processes for CEF, each cell conceals 7 bits(this is the key element of our approach).

The proposed method will hide two or three information bits ether, “01”, “10”,”00”,”11”, or, “000”, “001”, “010”, “011”, “100”, “101”, “110”, “111”, in each RGB Font Color Formatting Schema.

Using these properties will make all the Font Color Schema with concealed bits that appears visually same but statically have different vales, this makes the proposed method hard to be detected by human vision.

The proposed Stenography Method has been implemented using MATLAB and is a high-performance language for technical computing. Matlab function is an easy to use, user interface function that guides a user through the process of either encoding & decoding a message into or from the CEF respectively. In this work, Matlab is implemented for processing LSB steganography technique with different CEF size, below is the snap shots of Fig. 9a-j.

Fig. 9 a : shows the embedded operation.

Fig. 9 a : shows the Hiding text operation.

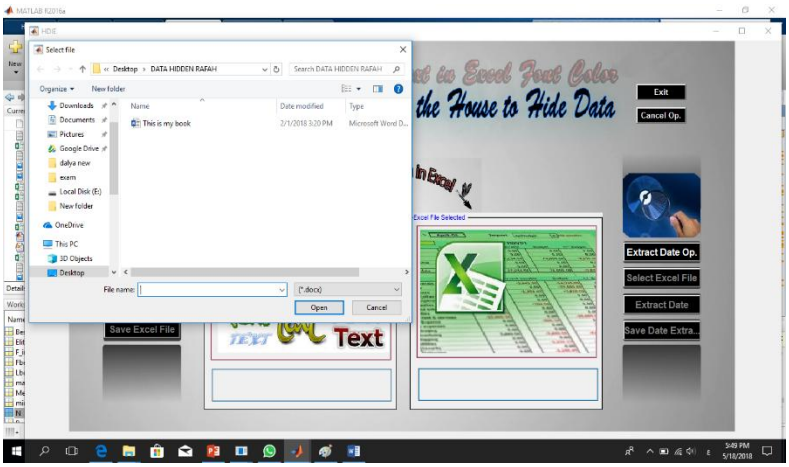


Fig. 9c: shows the selecting word file for text hiding operation.

Fig. 9d: shows the selecting CEF for embedding operation.

Fig. 9e: shows the ending of embedding process by saving the CEF.

Fig. 9f: shows the starting of extracting process.

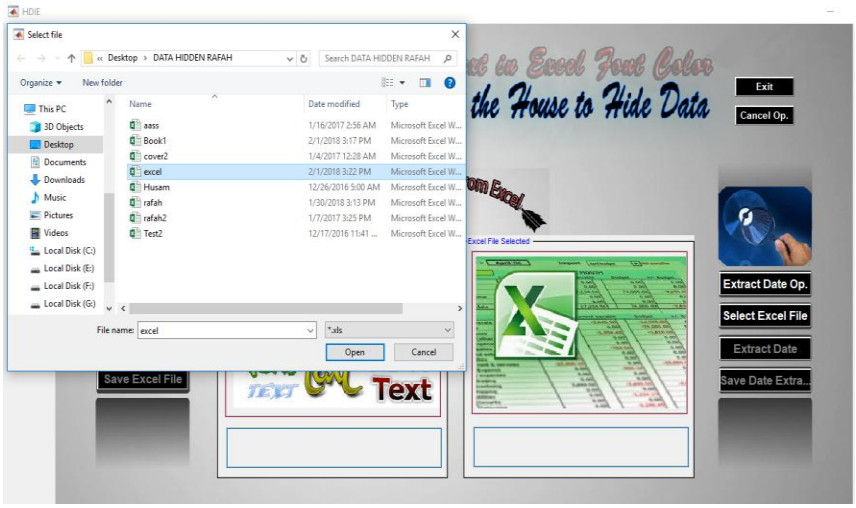


Fig. 9g: shows the selecting of CEF of extracting process.

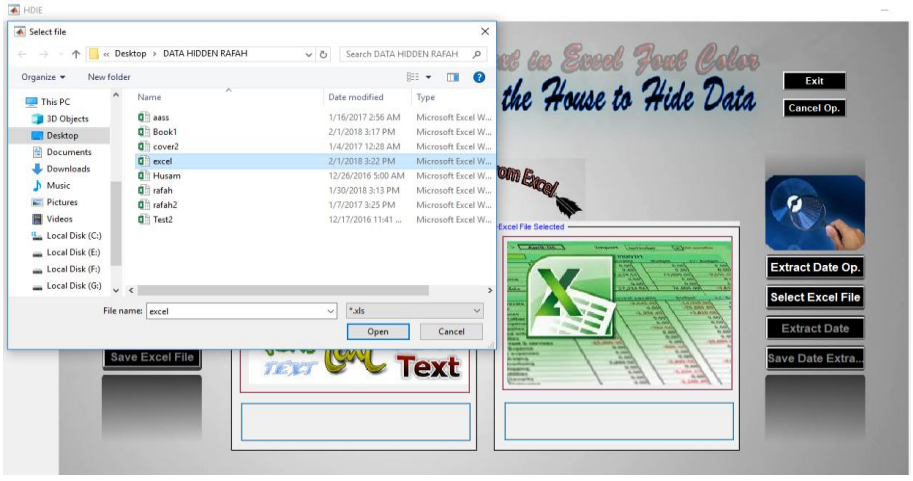


Fig. 9h: shows the selecting CEF from the determined folder.

Fig. 9i : shows the extraction text from the CEF.

Fig. 9j: shows the ending of extracting process and saving operation.

Hiding (embedding process) part: as shown in the Fig. 10, this part has the following steps.

Step 1: create an excel file called CEF and fill it with plain text or cover message. Otherwise find the suitable CEF document that is previously done and filled with the cover message.

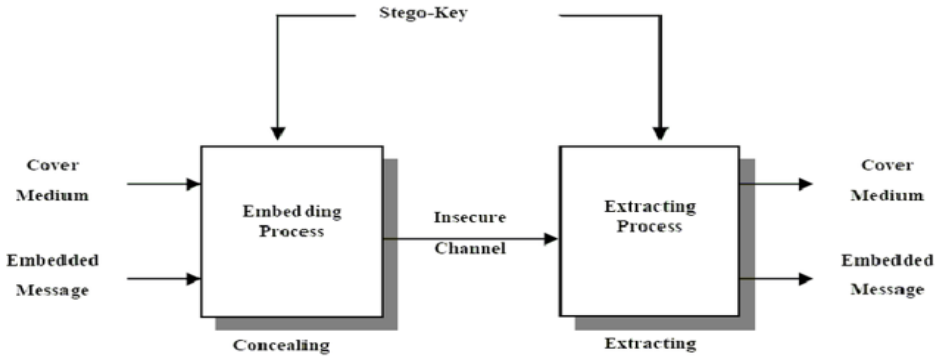


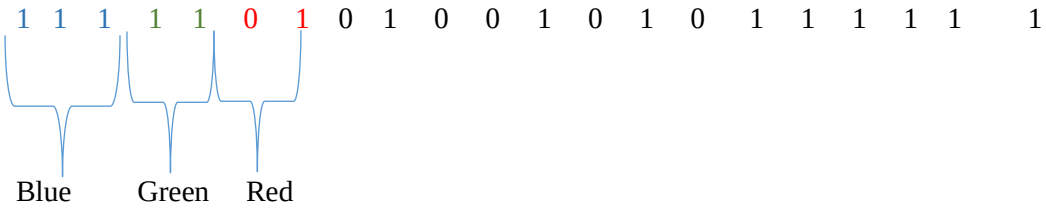
Fig. 10: the general paradigm for embedding and extraction stenography process

- Step 2 : find the first cell according to its turn using the auto-program as shown in the Figure. 8 along to the psedo-code algorithm-2, and then fill the first cell with bits coding (7, 6, 8 bits coding according to the max character presentation), fill the 2^{nd} cell till 5^{th} with the length of the message stream required during the embedding process, 6^{th} cell fill with formatting schema(in our example would be 1 as font color formatting).
- Step 3 : coded the secret message (each character in 7 bits or more according to ASCII code table, shown in Fig. 3) till the last characters in the secret message.
- Step 4 : Apply the columnar, concatenation transfer, and flipping operation for the secret message using bits coding stream, as shown in the Fig. 6a-b, the results of the

preprocessing now applicable to scatter into CEF by cell's turn as shown in the Fig. 8 along to psedo-code algorithm-2, this step will be repeated till the last character in the secret message, eventually code stream is ready to scatter into CEF for our example secret message (MSG) shown in Fig. 11.

1. Select the none-empty cell and prepare it for hiding process, according to the cell's turn as shown in Fig. 8 along to psedo-code algorithm-2.
2. Embedding the 7 bits coding stream from the flipping Matrix, first two bits into Red Color Font Formatting, Second Two bits into Green Color Font Formatting, Last three into Blue Color Formatting.
3. Save the final excel file documents (that containing the plain text and invisible secret message).

(noted the step 4 would be repeated till the code stream is over)



Color formatting for none- empty cell of CEF

Fig. 11 Secret Bits order in the process

Extracting (Retrieval) Part: as shown in Fig. 10, this part has the following steps:

- Step 1: open the corresponding excel documents which is including the plain text along with the secret message.
- Step 2: According to the manner for hiding the bits that shown in Fig. 8, first of all, the first cell is reserved for bits coding, the 2^{nd} till the 5^{th} are reserved for the length of secret message, while the 6^{th} cell is reserved for color Formatting Schema, after retrieving the header information that will be the guideway to extract the hidden message.
- Step 3: Continue fetching the none-empty cells in order to extracting the secret bits from the corresponding cell as the following, read the first two bits from Red Color Font Formatting, second two bits from Green Color Font Formatting, last three bits from Blue Color Font Formatting, carry on retrieving the next cell (Step 3 will be repeated until character length counter is over).
- Step 4: Retrieve the Secret Message from the coded stream.

References

[1] WesamBhayaetal, "Text Steganography Based on Font Type in MS-Word Documents", Journal of Computer Science 9 (7),PP. 898-904, 2013.

[2] Por, L.Y., K.S. Wong and K.O. Chee, "A Text-Based Data Hiding Method Using Unicode Space Characters", J. Syst. Soft., 85: PP. 1075-1082, 2012.

[3] Khandekar, S.A. and M.R. Dixit, "Steganography for Text Messages Using Image", J. Elect. Commun. Eng., 2:PP. 01-04, 2012.

[4]Saber, A. S. and W.A. Awadh, "Steganography in MS Excel Document Using Unicode System Characteristics", J. Basrah Res. Sci., 39: PP.10-1 9,2013.

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

[6] Por, L.Y. and B. Delina, " Information Hiding: A New Approach in Text Steganography" , Proceedings of the 7th WSEAS International Conference onApplied Computer and Applied Computational Science, Apr. 6-8, Hangzhou, China, PP. 1-7, 2008.

[7]Md. Khairullah, "A Novel Text Steganography System Using Font Color of the Invisible Characters in Microsoft Word Documents", December 28 - 30, 2009 IEEE Computer Society Washington, DC, USA ©2009, ICCEE '09 Proceedings of the 2009 Second International Conference on Computer and Electrical Engineering - Volume 01.

[8]Samraa Al-Asadi and Wesam S. Bhaya, "Text Steganography in Excel Documents Using Color and Type of Fonts" ,Research Journal of Applied Sciences 11(10):PP.1054-1059, January 2016.

[9]Bin Yangetal"Steganography in Ms Excel Document using Text-rotation Technique ", Information Technology Journal Volume 10 (4): PP.889-893, 2011.

[10]Gutub, A. and M. Fattani, " 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, 2007, Vienna, Austria, PP. 28-31, 2007.

[11]Provos, N. and P. Honeyman, "Hide and Seek: An Introduction to Steganography", IEEE Secur. Privacy, 1:PP. 32-44, 2003.

[12]Chen, B. and G.W. Wornell, "Quantization Index Modulation: A Class of Provably Good Methods for Digital Watermarking and Information Embedding", IEEE Trans. Inform. Theory, 47: PP. 1423-1443, 2001.