Novel Content Independent Steganographic Method for Microsoft Office

Saqib Ishtiaq, Aihab Khan and Basheer Ahmad Samim

Abstract— **Microsoft Office is widely used suit in offices, homes and educational institution. Some methods exist to hide the secret message by using MS Office software. Existing methodologies are depending on the contents of files to hide secret message. So these methods have low capacity as numbers of bit to be hidden are limited to the length of contents. In this paper, a novel, content independent method is proposed to hide secret message in MS Office. For this purpose several properties of object model of MS Word, MS Excel and MS Access are used. These properties have no relation to contents so this method is robust against attacks. Experimental results show that the proposed method has high embedding capacity as numbers of object added are limited to the available memory. By using this method there is no change in appearance of the files so it is imperceptible. This method shows small overhead on the size of files.**

***Index Terms*— MS Office Steganography; Object Model; Robust Steganography; High Capacity Steganography**

I. Introduction

Microsoft Word, Excel and Access are part of Microsoft Office and powerful word processor, spreadsheet and database software respectively. Each Microsoft Office application has object model which is used for automation and interaction [1]. Objects in object model are arranged in hierarchal manner [1], [2]. All the objects in object models have several properties in which some of them are read only while other have read/write permission. We can update these properties by using Visual Basic for Application (VBA) or Visual Studio Tool for Office (VSTO). Most of the properties of the object models have direct relation with the contents of the files. Therefore any change in its properties also affects the contents of the files. As some of properties are content independent. Manipulation with these properties has no effect on the contents of the files. These properties are used for the steganography.

Steganography is technique to hide secrete message in the cover medium (text, audio and video) without being into the knowledge of eavesdroppers [3]. Steganography is used for the covert communication between the parties. There are several methods used to conceal the secret message in the MS Word and Excel files. Most of the existing methods for concealing the secret message in MS Office suite use MS Word Documents. Existing methods can be categorized on the bases of properties [1], [2], format [4]-[6], [8], [11], etc. [3], [7], [9].

Properties of the MS Word object model are used to embed the bits to the range of the documents [1], [2]. Firstly message is converted to binary and then layout of document is transformed according to the bits. These methods are not robust against the attacks. RGB color based method is proposed in [4]. Firstly message is converted to the eight bits binary stream. RGB color values of imperceptible characters are changed according to the binary stream. Mahato *et al.* [5] proposed a method to hide secret message by changing the size of space character. Stojanov et al. [6] proposed four formats based methods hide message. In first method character size, in second method character underlined, in the third paragraph border and in the last method sentence border are used to embed the secret message. Bhaya et al. [8] proposed a method to secret message by using same Font types. Change the case of the character in MS Word document to capital is used to hide bit 1 and left it small to embed bit 0 [11]. Ray *et al.* [12] proposed a method which encrypt the secrete message before the embedding. Cipher is converted to binary form. To hide a secrete bit blank space is selected randomly. Selected space is replaced with the ASCII code 160 to hide bit 1 and left unchanged to hide bit 0. By using this method maximum 1 bit can be hidden per blank space in the ASCII text. Change tracking feature of MS word is used to hide the secrete data in the document. Segment of document is degenerated during the embedding process then document is converted back to the original form and changes made are tracked [13].

Khairullah *et al.* [3] proposed a method to concatenate arbitrary number of zeros in the start of number or after the decimal point in the end of number. In this way maximum two bits can be hidden in the one number in financial statement. This approach has low capacity as there are few locations in which bits zeros can be padded. This approach is also not imperceptible as extra zeros are visible in the statement. Text in the MS Excel workbook cells can be used to hide the secret bits. To hide the bit 1 angle of the text is rotated and remains unchanged for bit 0 [9].Saber et al. proposed Unicode based method to hide messages [10]. Alternate Unicode is used to hide bit 1 and remains unchanged to hide bit 0.

No considerable work has been found as per our knowledge to use MS Access database for embedding the secret message. There is robustness issue with [1]-[6], [8]-[11] as message is hidden in the text. Deletion and replacement attack damages the message. Message hidden by format based methods can be destroyed by format attacks. There is imperceptibility issue

Manuscript submitted on November 17, 2016.

Saqib Ishtiaq is with the Iqra University Islamabad, Pakistan (Phone: 0092-313-5477801; e-mail: darkalishershahi@gmail.com).

Aihab Khan, is with Department of Computing, Iqra University Islamabad, Pakistan (e-mail: aihabkhan@yahoo.com).

Bashir Ahmad Samim is with Management Science Department, Iqra University Islamabad, Pakistan (e-mail: drbashir@iqraisb.edu.pk).

with [2], [3]. Method [3], [5], [9] have low capacity as few bits can be hidden by these methods.

In this paper a novel content independent method is proposed to conceal the message. Content independent properties are used to hide the message without affecting the contents of covered files. Existing methods depends on the contents of files to hide the secrete message so presence of contents in the files is essential. Message cannot be concealed in blank files by using existing methods. In proposed method content independent properties are used instead of contents so message can also be hidden in blank files.

II. PROPOSED METHOD

In this paper we propose novel method to hide secrete message in object model of MS Word document, MS Excel Workbook and MS Access Database. During the embedding of message .docx format of MS Word document, .xlsx format of MS Excel workbook and .accdb format of MS Access database are used. While secrete message is in form of ASCII text. Variable, Name and Properties objects are content independent objects. Fig.1. Show objects and properties used to conceal data. Variable is part of MS Word Document and used to store macro settings among different session [1]. Name object is part of MS Excel Object model. It represents the name for the cell range. Properties collection objects of MS Access Database contain all the properties. All these objects have properties and methods. Properties of these objects are used to embed secret message. We use Name and Value properties of Variable object of Microsoft Word, Name, referTo, Comments and Visible properties of Name object of the Microsoft Excel and Name and Value property of properties object of Microsoft Access. Name properties require valid name so spaces and symbols cannot be included so one word is assigned to name property. Name must be unique so to maintain the uniqueness, object number is appended at the end. For referTo property each character is converted to the ASCII value and then cell reference is calculated which is assigned to referTo property. Visible property is used to make the names invisible. Comments and value properties can hold 255 characters. Secrete message is in ASCII text format. Fig. 2 shows the model for proposed method. For embedding secret message MS Word Document, MS Excel Workbook and MS Access Database are used. Object and property are selected randomly. Suitability of property is evaluated for embedding message. Secret Message is embedded to the property.

In our method there are two separate algorithms. Embedding algorithm is used to embed the message while retrieving algorithm is used to retrieve the message from the carrier files.

A. Embedding Algorithm

```
Input: Secrete Message, MS Word Document, MS Excel
Workbook, MS Access Database
Output:  Covered MS Word Document, MS Excel Workbook,
MS Access Database
Step 1: random (object)
Step 2: random (property)
Step 3: if (suitability for embedding) = True Then
```

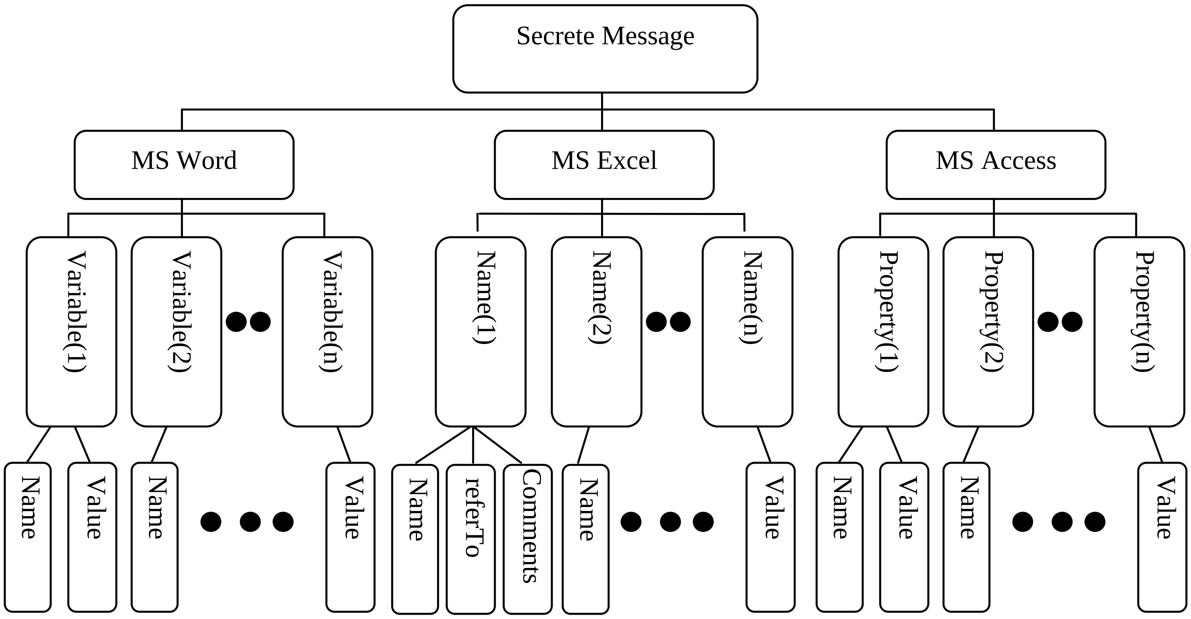


Fig. 1. Objects and properties used

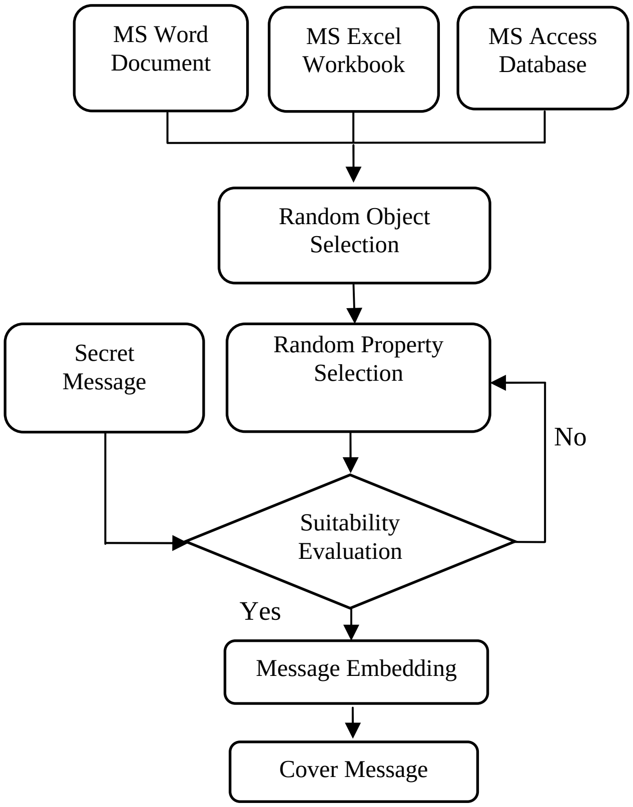


Fig. 2. Model for proposed method

```
        Embed(message)
        Else
          Goto Step 2
Step 4: If (NOT End of Secret Message) Then Goto Step 1
Step 5: Commit Changes
```

B. *Retrieving Algorithm*

```
Input: Covered MS Word Document, MS Excel Workbook,
MS Access Database
Output: Message
Step 1: Random (Object)
Step 2: Random (Property)
Step 3: If valid then
          propertyString=toString(property)
          else
          Goto Step 2
Step 4: Message=concat(Message,propertyString)
Step 5: If  Not End of properties Goto step 1
```

III. Results and Discussion

In this research embedding capacity, imperceptibility, robustness and overhead on file size is get improved.

A. *Capacity Analysis:*

Existing methodologies depend on the contents of files to embed the secret message. Capacity of cover medium is improved with increase in size of contents. Our method has high capacity as we can add number of objects is limited to available memory of the system. There is no need of contents in files to hide the secret message therefore message can be added to the blank MS Word Document, MS Excel Workbook and MS Access database. This method shows very little overhead on the size of files. Two sets of files are used for the experiments as shown in Table I.

TABLE I
FILE SIZE OF VARIOUS FILES USED

Files	MS Word Document (byte)	MS Excel Workbook (byte)	MS Access Database (byte)
Blank	9,821	7,822	286,720
Non Blank	1,239,040	643,257	2,707,456

Messages of different volume are embedded and extracted successfully. CIS (Change in size), PCIS (Percentage Change In Size) and CISTNECR (Change In Size To Number of Embedded Character Ratio) are computed. Following formula is used.

Change in size (CIS)=After Embedding Size-Original Size

$$\text{PCIS}=\frac{CIS}{Original\ Size}\times 100\%$$

$$\text{CISTNECR}=\frac{CIS}{Number\ of\ Embedded\ Character}$$

$$NECTCISR=\frac{Number\ of\ Embedded\ Character}{CIS}$$

Smaller values for the CIS, PCIS and CISTNECR are desirable. While larger value for NECTCISR shows indicates better results. Experimental results of blank and non blank files are shown in the Table II and Table III respectively. Table II shows the experimental results after embedding secret message to the blank document. Table III shows the experimental result on the non blank files. As per best of our knowledge no steganographic method exists to embed secret message to the blank files. Existing methods are not applicable on blank files so comparison is given only for non blank files. As depicted by the results CIS, PCIS and CISTNECR are decreased for the larger file. Comparison of CIS, CISTNECR with existing methods is shown in Table IV and V respectively. Proposed method has low CIS and CISTNECR values as compared to the existing methods [6]. So the proposed method shows little impact on size of carrier files. For small message there is no change in size.

B. *Imperceptibility Test*

This method is imperceptible as message is embedded invisibly. Embedded message cannot be seen by the human eye and cannot be heard by the human ear, to retrieve the message from the carrier files a special program is required. During the embedding secrete message "Iqra University" is concealed in the files. After embedding the secrete message in the files there is no change in the appearance of the files. Before embedding and after embedding appearance of the files are shown in fig. 3 and fig. 4 respectively.

TABLE II
EXPERIMENT RESULTS FOR BLANK FILES

Number of Embedded Characters	AES (byte)	CIS (byte)	PCIS	CISTNECR	NECTCISR
100	304,363	0	0	-	-
500	304,576	213	0.069982225	0.426	2.34741784
1000	304,787	424	0.13930734	0.424	2.358490566
5000	314,655	10,292	3.381488551	2.0584	0.485814225

TABLE III
EXPERIMENT RESULTS FOR NON BLANK FILES

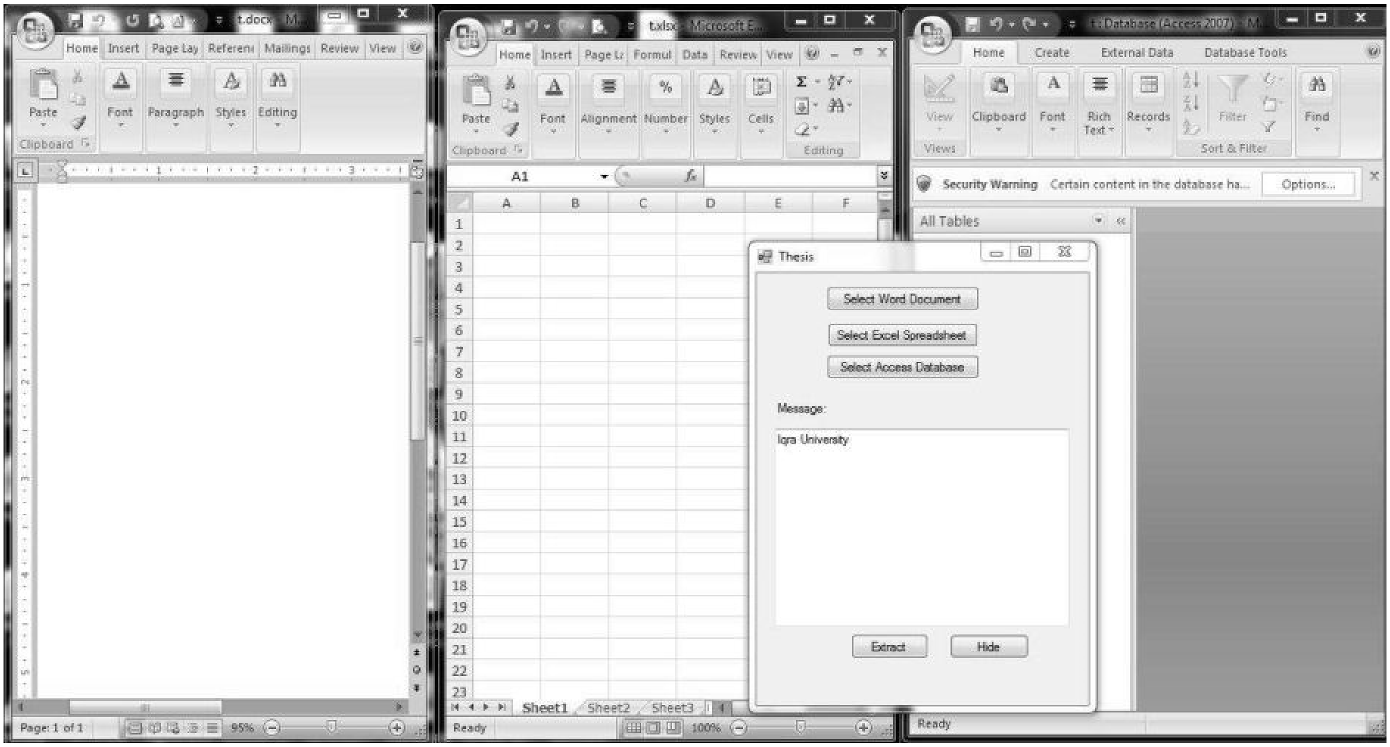
Number of Embedded Characters	AES (byte)	CIS (byte)	PCIS	CISTNECR	NECTCISR
100	4,589,753	0	0	0	-
500	4,589,966	213	0.004640773	0.426	2.34741784
1000	4,590,471	718	0.015643543	0.718	1.39275766
5000	4,593,599	3,846	0.083795359	0.7692	1.300052002

TABLE IV
COMPARISON OF CHANGE IN SIZE (IN BYTES) WITH EXISTING METHODS [6].

Number of Embedded Characters	Proposed	Character Scale	Character Underline	Paragraph Borders	Sentence Border
100	0	1,715	262	285	2,466
500	213	5,689	2,781	1,919	6,547
1000	718	16,058	6,470	4,131	14,646
5000	3846	92,973	18,765	NA	NA

TABLE V
COMPARISON OF CISTNECR WITH EXISTING METHODS [6].

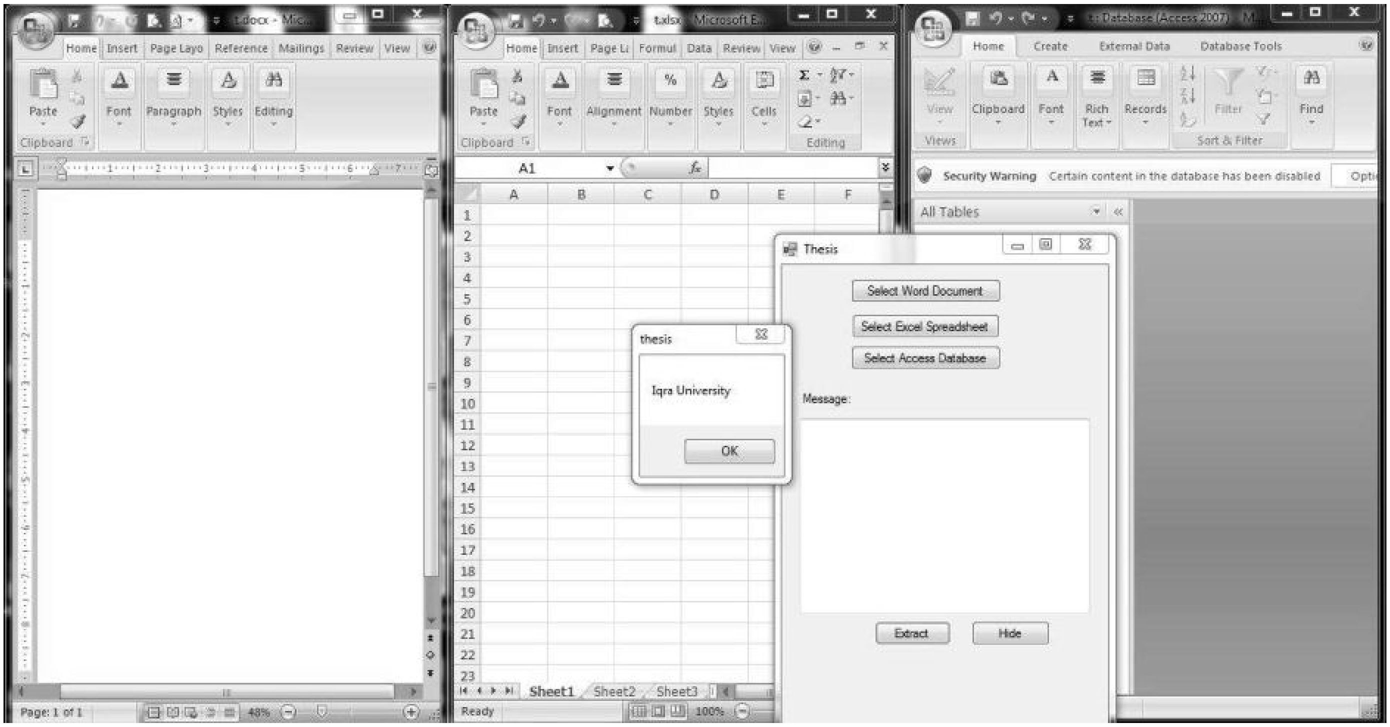
Number of Embedded Characters	Proposed	Character Scale	Character Underline	Paragraph Borders	Sentence Border
100	0	17.15	2.62	2.85	24.66
500	0.426	11.378	5.562	3.838	13.094
1000	0.718	16.058	6.47	4.131	14.646
5000	0.7692	18.5946	3.753	NA	NA



Security Warning Certain content in the database ha... Options...

Iqra University

Fig. 3. Before embedding the secret message blank files



Security Warning Certain content in the database has been disabled

Iqra University

Fig. 4. After embedding blank files and retrieved message

C. *Robustness Test*

Robustness is capability of resistance against the modification or elimination of the concealed data [14]. In steganography main aims to protect the hidden message from alteration or elimination. If message is retrieved successfully without any modifications after the attacks then the method is said to be robust. Proposed method is robust against the attacks as after the insertion, deletion, replacement and format attack the message is retrieved successfully as the secrete message is present in the objects rather than the contents of files . The message remains intact in the objects after the attacks and extracted without modification. So message concealed by the proposed method cannot be changed or removed from the covered files.

IV. Conclusion

Microsoft Word, Microsoft Excel and Microsoft Access have their object models and each object has properties and methods. Some properties of the object models can be used to store strings. These properties can be used to hold secrete message. By hiding messages by using these properties there is no need of contents in files for embedding. This method shows high capacity with low overhead. This method is more robust as after attacks message is retrieved successfully.

References

[1]. Zhang, Y., Qin, H., & Kong, T. (2010). A novel robust text watermarking for word document. 2010 3rd International Congress on Image and Signal Processing. Yantai, China . 38-42

[2]. Khadim, U., Khan, A., Ahmad, B., & Khan, A. (2015). Information Hiding in Text to Improve Performance for Word Document. International Journal of Technology and Research, 3(3), 50-55.

[3]. Khairullah, M. (2014). A Novel Text Steganography System in Financial Statements. IJDTA International Journal of Database Theory and Application, 7(5), 123-132.

[4]. Khairullah, M. (2009). A Novel Text Steganography System Using Font Color of the Invisible Characters in Microsoft Word Documents. Second International Conference on Computer and Electrical Engineering. Dubai . 482–484

[5]. Mahato, S., Yadav, D. K., & Khan, D. A. (2013). A Novel Approach to Text Steganography Using Font Size of Invisible Space Characters in Microsoft Word Document. Intelligent Computing, Networking, and Informatics Advances in Intelligent Systems and Computing,243, 1047-1054.

[6]. Stojanov, I., Mileva, A., & Stojanovic, I. (2014). A New Property Coding in Text Steganography of Microsoft Word Documents. The Eighth International Conference on Emerging Security Information, Systems and Technologies). Lisbon, Portugal. 25-30.

[7]. Kingslin, S., & Kavitha, N. (2015). Evaluative Approach towards Text Steganographic Techniques. Indian Journal of Science and Technology, 8(29), 1–8.

[8]. Bhaya, W., Rahma, A. M., & Al-Nasrawi, D. (2013). Text Steganography Based On Font Type In Ms-Word Documents. Journal of Computer Science, 9(7), 898-904.

[9]. Yang, B., Sun, X., Xiang, L., Ruan, Z., & Wu, R. (2011). Steganography in Ms Excel Document using Text-rotation Technique. Information Technology Journal, 10(4), 889-893.

[10]. Saber, A. S., & Awadh, W. A. (2013). Steganography in MS Excel Document Using Unicode System Characteristics. Journal of Basrah Researches ((Sciences)) Vol, 39(1),10-19.

[11]. Aminali, A., & Saad, A. S. (2013). New Text Steganography Technique by using Mixed-Case Font. International Journal of Computer Applications, 62(3), 6-9

[12]. Ray, R., Sanyal, J., Das, D., & Nath, A. (2012). A new Challenge of hiding any encrypted secret message inside any Text/ASCII file or in MS word file: RJDA Algorithm. In *Communication Systems and Network Technologies (CSNT), 2012 International Conference*, 889-893.

[13]. Liu, T. Y., & Tsai, W. H. (2007). A new steganographic method for data hiding in microsoft word documents by a change tracking technique. *IEEE Transactions on Information Forensics and Security*, *2*(1), 24-30.

[14]. Gutub, A., & Fattani, M. (2007). A novel Arabic text steganography method using letter points and extensions. In *WASET International Conference on Computer, Information and Systems Science and Engineering* , 25-27.