Abdul Monem S. Rahma, Wesam S.Bhaya, Dhamyaa A. Al-Nasrawi / International Journal of Engineering Research and Applications (IJERA) ISSN: 2248-9622
www.ijera.com Vol. 3, Issue 4, Jul-Aug 2013, pp.1153-1165

Text Steganography Based On Unicode of Characters in Multilingual

Prof. Abdul Monem S. Rahma[1], Wesam S.Bhaya[2], Dhamyaa A. Al-Nasrawi[3]

[1](Computer Science Department, University of Technology, Iraq
[2] (College of Computer Technology, University of Babylon, Iraq
[3](Computer Science Department, Kerbala University, Iraq

ABSTRACT

Network security, and secure communications through public and private channels are more important issue specially when computer usage is increasing, for both social and business areas. Data hiding is one of approach to obtain a secure communication medium and protecting the information during transmission. Text steganography is most challenging because of the presence of very less redundant information in text documents as compared to the images and audio. In this paper a novel method is proposed for data hiding in English scripts using Unicode of English alphabet in another languages. In this method, 13 characters from English alphabet was chosen for hiding process which have appearance in another languages. Two bits embedded in one time, using ASCII code for embedding 00, and using Unicode of multilingual for embedding 01, 10, and 11. This method has a height hiding capacity based on specific characters in each document. As well as have very good perceptual transparency and no changes in original text.

Keywords- Multilingual characters, Text hiding, Unicode standard.

I. Introduction

One way of secure data transfer over the Internet is steganography, which conceals the existence of a message [1]. When information hiding is used, even if an eavesdropper snoops the transmitted object, he cannot surmise the communication since it is carried out in a concealed way. Steganography overcomes the limitation of cryptography(that the third party is always aware of the communication because of the unintelligible nature of the text) by hiding message in an innocent looking object called cover[2].

In modern steganography use electronic media rather than physical objects and texts. The text to be concealed is called embedded data. An innocuous medium, such as text, image, audio, or video file; which is used to hide embedded data is called cover. The stego object is an object we get after hiding the embedded data in a cover medium[1].

There are a number of researches had already explored in new steganographic techniques in texts, such as white spaces [3], Synonyms[4], Word Shifting [5], and Line shifting [6]. This paper focused on researches which used Unicode in data hiding, M. H. Shirali-Shahreza, and Mohammad Shirali-Shahreza proposed new method for hiding \ information in Persian and Arabic Unicode texts [7]. Also, they proposed another method for hiding data in Persian (Farsi) and Arabic texts. They based on characters of « ی» and « ک» have the same shape but different codes [8]. Lip Yee Por and et al. proposed method based on Unicode space characters with respect to embedding efficiency [9].

In this paper, some letters of English alphabet have appearance in another languages with different codes, these letters used in the proposed method for data hiding in English scripts using Unicode of English alphabet in another languages(multilingual).

The rest of the paper is organized as follows. Section 2 explain the Unicode Standard. Section 3 presents the proposed method in detail. Section 4 and 5 demonstrate the results and concludes of the proposed method.

II. Unicode Standard

The Unicode Standard is the universal character encoding scheme for written characters and text. It defines a consistent way of encoding multilingual text that enables the exchange of text data internationally and creates the foundation for global software[10]. Unicode can be implemented by different character encodings. The most commonly used encodings are UTF-8, UTF-16. UTF-8 uses one byte for any ASCII characters, which have the same code values in both UTF-8 and ASCII encoding, and up to four bytes for other characters. UCS-2 uses a 16-bit code unit (two 8-bit bytes) for each character[11].

Unicode characters are distinguished by code points, which are conventionally represented by the letter U followed by four or five hexadecimal digits, for example U+00AE or U+1D310. Unicode characters can range in scalar values from 0 to over a million. The entire range of Unicode characters is divided into 17 blocks, each block is referred to as a plane and is numbered starting from 0. Characters in the Basic Multilingual Plane (BMP), containing

modern scripts – including many Chinese and Japanese characters – and many symbols, have a 4-digit code. Historic scripts, but also many modern symbols and pictographs (such as emoticons, many CJK characters, and Egyptian Hieroglyphics) have 5-digit codes[11].

Then, Unicode refer to the family of standards and technologies associated with the Unicode Consortium that can be utilized for working with a written language in a computer environment[12].

III. Proposed Method

In this paper, a new method was presented for text steganography in English scripts using Unicode of multilingual characters. Many English alphabet characters (called Latin alphabet) have a good appearance in another languages, with different codes and different glyphs, see Table 1.

Unfortunately, not all of these characters can be used in hiding process, because the glyphs of them dissimilar to original English scripts. Just 13 characters was chosen for hiding process based on the following criteria:

1. Located in Plane 0: which contains U+0000 - U+FFFF, this is known as the Basic Multilingual Plane (BMP).
2. The glyphs: A ***glyph*** is a presentation of a particular shape which a character may have when rendered or displayed. The glyph of selected characters must be similar to the original script.
3. Availability: depends on its presence in the specified font (selected characters must be supported by standard fonts)[11]. If a desired character is not present in the available fonts(that mean characters have no glyph), an empty box, a question mark or another replacement will be shown: **?**, see Table 1 .

Table 2 explain the selected characters in hiding process.

In proposed method, two processes was implemented, hiding process, and extracting process. Hiding process based on the appearance of selected characters in english script. In this method two bits can be embedding at one time. Firstly, selected characters must to be found in document, then embedding process implemented by replacement depending on secret message which can be hidden. and replacement can be done based on secret message as follows:

$$\text{If secret message} = \begin{cases} 00 & \text{No change} \\ \{01,10,11\} & \text{Replace with nicode of Multilingual characters based on Table 2} \end{cases} \quad \ldots(1)$$

The document file must be has enough area to hide data (secret message with two bytes represent message length embedded in beginning). This process called Checking capacity of hiding , see Fig.1.

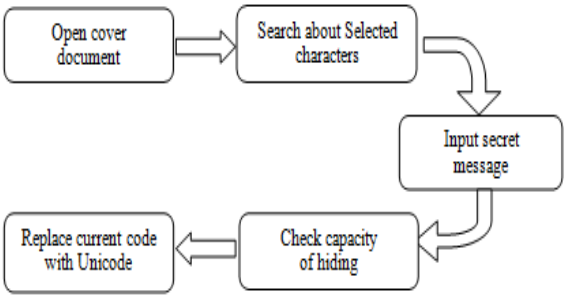

Figure 1. Hiding process

Hiding process (Embedding process) summarized in the following algorithm:

Embedding algorithm
Input : Document file, and secret message.
Output : Stego file.
1. Open cover document.
2. Scan cover document to find selected characters in Table 2,
3. Compute number of selected characters to check the capacity of hiding.
4. Get binary form of secret message.
5. For each two symbol in secret message
   - if bit = 00, then no change (ASCII code) , else replace by Unicode of Multilingual characters in Table 2.
6. Hide the message length in the beginning of secret message.
7. Return stego document.

English scripts written in (Latin letters), take the range (U+0041-U+005A) for uppercase Latin alphabet, and (U+0061-U+007A) for lowercase Latin alphabet in hexadecimal. For example, there is Latin capital "A" which is defined (U+0041) can be no change if secret message is 00, but can be replaced by one of Unicode of Multilingual characters in Table 2 , Based on secret message, replaced by (U+0391) when secret message is 01 as example.

In another side, the extracting process Fig.2 is the opposite operation of the hiding process.

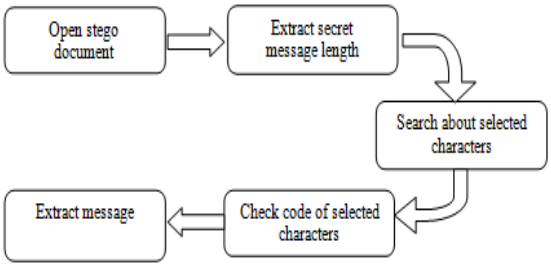

Figure 2. Extracting process

In order to find selected characters, we must check the code of current characters, by the following statement:

If Code of Selected characters
- In range of (U+0041 - U+005A) or (U+0061 - U+007A), then Secret message = 0 …(2)
- Else Secret message = {01,10,11} based on Table 2

The first 16-characters represent the secret message length. Extracting process summarized in the following algorithm:

Extracting algorithm
Input : Stego file .
Output : Secret message.
1. Open stego document.
2. Extract secret message length
3. Scan stego document to find ligatures characters,
4. Check code of selected characters based on equation (2)
   - if code in uppercase or lowercase English Latin letters then secret message=00
   - else secret message=01,or 10, or 11 based on Table 2.

IV. Results

In this paper, the proposed method of data hiding is tested by taking different cover documents with different sizes and hiding the same secret message in some of them, sees the corresponding GUI for the proposed method in Fig. 3.

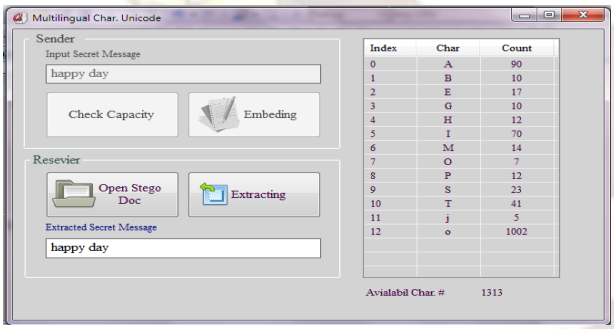

Figure 3. GUI of proposed method

The payload of bits can be hidden based on selected characters number. If the cover file contains, for example, 200 selected characters, we can hiding 400 bits in it(because two bits can be embedded in one character). Fig. 4 and Fig. 5 represent cover document and stego document respectively.

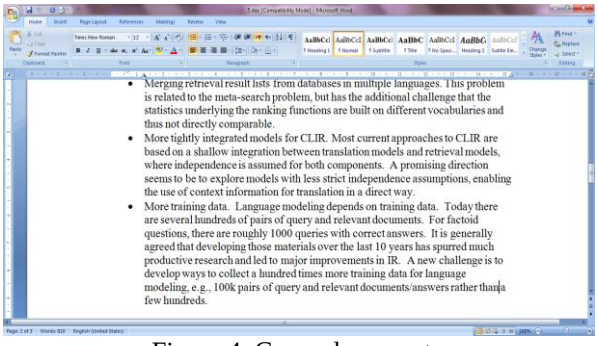

Figure 4. Cover document

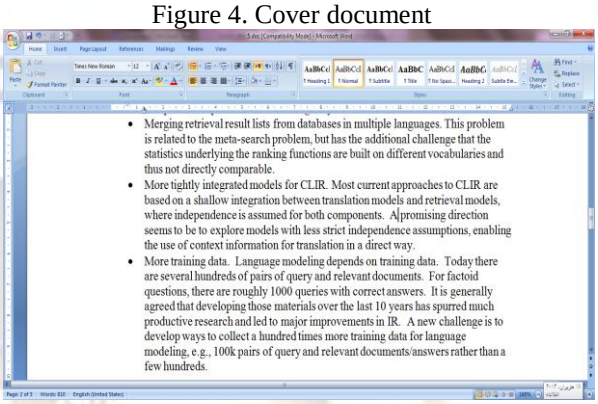

Figure 5. Stego document

The results that are got from these experiments can be summarized in the Table 3. Note the average of size for stego document (after hiding) is increased about (11.1 %) from original size, because using Unicode instead of ASCII code in embedding process, which use UTF-16 encodings for each character.

V. Conclusions

In data hiding method, the main goals of steganography are (perceptual transparency, capacity, and robustness). Proposed method has an excellent perceptual transparency because the stego text is similar to the original text using characters in multilingual which their glyphs more similar and suitable to glyphs of original script. The hiding capacity of proposed method is very high, depending on selected characters frequency shown in Table 2. In addition, this method is robust to digital copy-past operation, which means that copying and pasting the text between computer programs preserve hidden information. In other side, there is an increasing in stego document size, result from Unicode which use 2-bytes for each character. Huffman code for data compression can be suggested to overcome this drawback. Thus, the proposed method provide an efficient text steganography method by Unicode, and an excellent way to obtain safe and secure information transformation.

References

[1] M. Agarwal, Application Text Steganographic Approaches: A Comparison,

International Journal of Network Security & Its Applications, 5(1), 2013, 91-106.

[2] K. Rabah, Steganography-the art of hiding data, *Information Technology Journal*, 3(3), 2004, 245-269.

[3] W. Bender, D. Gruhl, N. Morimoto, and A. Lu, *Techniques for data hiding, IBM System Journal 35(3-4),* 1996, 313-336.

[4] H. Shirali-Shahreza, M. Shirali-Shahreza, A new synonym text steganography. *Proc. 8th International Conf. on Intelligent Information Hiding and Multimedia Signal Processing*, Harbin, China, 2008, 1524-1526.

[5] Y. Kim, K. Moon, I. Oh, A text watermarking algorithm based on word classification and inter-word space statistics, *Proc. 7th International Conf. on Document Analysis and Recognition*, Edinburgh, Scotland: IEEE,2003, 775-779.

[6] S. Low, N. Maxemchuk, J. Brassil, L. O'Gorman, Document marking and identification using both line and word shifting, *Proc.14th Annual Joint Conf. of the IEEE Computer and Communications Societies*, Boston, MA, USA: IEEE, 1995, 853- 860.

[7] H. Shirali-Shahreza, M. Shirali-Shahreza, Steganography in Persian and Arabic unicode texts using pseudo-space and pseudo-connection characters, *Journal of Theoretical and Applied Information Technology, 4(8),* 2008, 682-687.

[8] H. Shirali-Shahreza, M. Shirali-Shahreza, Arabic/Persian text steganography utilizing similar letters with different codes, *The Arabian Journal for Science and Engineering, 35(1B),* 2010, 213-222.

[9] L. Por, K. Wong, K. Chee, UniSpaCh: A text-based data hiding method using unicode space characters, *The Journal of Systems and Software 85*(5), 2012, 1075– 1082.

[10] The Unicode Consortium, The Unicode Standard, http://www.unicode.org.

[11] J. Korpela, *Unicode explained* (United States of America., O'Reilly Media , 2006).

[12] D. Yacob, Unicode for Under-Resourced Languages, 2006, 33-38. From http://mt-archive.info/LREC-2006-Jacob.pdf

.

Table 1: Unicode of Characters in Multilingual

(ASCII) Latin Alphabets	Multilingual, Glyphs, and Unicode						
C0 Control and Basic Latin	*Greek and Coptic*	*Cyrillic*	*Cherokee*	*Cherokee*	*Unified Canadian Aboriginal Syllabics*	*Unified Canadian Aboriginal Syllabics*	*Unified Canadian Aboriginal Syllabics*
A	A	A	A	Ꭿ	A	Ꭺ	Ꭰ
0041	0391	0410	13AA	13AF	15C5	15E9	15CB
	Phonetic Extensions	*Phonetic Extensions*	*Halfwidth and Fullwidth Form*				
	ᴀ	ᴬ	A				
	1D00	1D2C	FF21				
	Latin Extended-B	*IPA Extensions*	*Greek and Coptic*	*Cyrillic*	*Cyrillic*	*Cherokee*	*Unified Canadian Aboriginal Syllabics*
B	Ɓ	ʙ	Β	в	В	Ᏼ	Ᏼ
0042	0181	0299	0392	0432	0412	13F4	15F7
	Vai	*Halfwidth and Fullwidth Form*	*Coptic*	*Phonetic Extensions*	*Letterlike symbols*	*Coptic*	
	ꕢ	B	ⲃ	ᵬ	ℬ	ⳁ	
	A557	FF22	2C83	1D2E	212C	2C82	
	Greek and Coptic	*Cyrillic*	*Georgian*	*Cherokee*	*Phonetic Extensions*	*Numbers Form*	*Halfwidth and Fullwidth Form*
C	Ϲ	С	Ⴋ	Ꮯ	ᴄ	Ⅽ	C
0043	03F9	0421	10BA	13DF	1D04	216D	FF23
	Mathematical Operators	*Kayah Li*	*Coptic*	*Yi Radicals*	*Yi Radicals*		
	∁	ꤋ	ⳇ	ꑖ	ꑀ		
	2201	A90D	2CA4	A49D	A49F		
	Latin Extended-B	*Cherokee*	*Unified Canadian Aboriginal Syllabics*	*Unified Canadian Aboriginal Syllabics*	*Phonetic Extensions*	*Phonetic Extensions*	*Numbers Form*
D	Ɗ	Ꭰ	Ꭰ	Ꭰ	ᴰ	ᴅ	D
0044	018A	13A0	15DE	15EA	1D30	1D05	216E
	Letterlike symbols	*Halfwidth and Fullwidth Form*					
	ⅅ	D					
	2145	FF24					
	Greek and Coptic	*Cyrillic*	*Cherokee*	*Phonetic Extensions*	*Tifinagh*	*Phonetic Extensions*	*Hangul Compatibility Jamo*
E	Ε	Е	Ꭼ	ᴱ	ⵄ	ᴇ	ㅌ
0045	0395	0415	13AC	1D31	2D39	1D07	314C

	Halfwidth and Fullwidth Form	*Vai*					
	E	?					
	FF25	A5CB					
F 0046	*Greek and Coptic*	*Letterlike symbols*	*Unified Canadian Aboriginal Syllabics*	*Halfwidth and Fullwidth Form*	*Latin Extended-D*		
	F	*F*	F	F	?		
	03DC	2131	15B4	FF26	A730		
G 0047	*IPA Extensions*	*Cyrillic Supplement*	*Cyrillic Supplement*	*Cherokee*	*Cherokee*	*Cherokee*	*Phonetic Extensions*
	ɢ	G	ɢ	G	G	G	ᴳ
	0262	050C	050D	13C0	13B6	13E9	1D33
	Halfwidth and Fullwidth Form						
	G						
	FF27						
H 0048	*IPA Extensions*	*Greek and Coptic*	*Cyrillic*	*Cyrillic*	*Cherokee*	*Unified Canadian Aboriginal Syllabics*	*Phonetic Extensions*
	ʜ	H	H	н	H	H	ᴴ
	029C	0397	041D	043D	13BB	157C	1D34
	Ethiopic	*Coptic*	*Coptic*	*Runic*	*Letterlike symbols*	*Halfwidth and Fullwidth Form*	*Yi Syllables*
	?	?	□	ᚺ	ℍ	H	ꀈ
	12D8	2C8E	2C8F	16BA	210D	FF28	A03F
	Phonetic Extensions	*Letterlike symbols*					
	ᴴ	*H*					
	1D78	210B					
I 0049	*IPA Extensions*	*Greek and Coptic*	*Cyrillic*	*Cyrillic*	*Cherokee*	*Phonetic Extensions*	*Numbers Form*
	ɪ	I	I	І	T	ᴵ	Ⅰ
	026A	0399	04C0	0406	13C6	1D35	2160
	Halfwidth and Fullwidth Form	*Halfwidth and Fullwidth Form*	*Halfwidth and Fullwidth Form*	*Coptic*	*Latin Extended-D*	*VAI*	*Yi Syllables*
	I	ɪ	ɪ	?	?	?	ꀤ
	FF29	FF6A	FF74	2C92	A7FE	A56F	A024
	Tifinagh	*Bopomofo Extended*					
	?	ㄧ					
	2D4A	31B2					
J 004A	*Cyrillic*	*Myanmar*	*Cherokee*	*Unified Canadian Aboriginal Syllabics*	*Phonetic Extensions*	*Phonetic Extensions*	*Halfwidth and Fullwidth Form*

	J		J	J	ᴶ	ᴊ	J
	0408	1042	13AB	148D	1D36	1D0A	FF2A
	Letterlike symbols						
	ℐ						
	2110						
K 004B	*Latin Extended-B*	*Greek and Coptic*	*Cyrillic*	*Cherokee*	*Phonetic Extensions*	*Phonetic Extensions*	*Letterlike symbols*
	K	K	K	K	ᴷ	ᴋ	K
	0198	039A	041A	13E6	1D37	1D0B	212A
	Halfwidth and Fullwidth Form	*Coptic*	*Runic*				
	K	□	ᛕ				
	FF2B	2C94	16D5				
L 004C	*IPA Extensions*	*Armenian*	*Armenian*	*Cherokee*	*Unified Canadian Aboriginal Syllabics*	*Tai Le*	*Phonetic Extensions*
	ʟ	Լ	l	L	ᒪ	ᥣ	ᴸ
	029F	053C	056C	13DE	14AA	1968	1D38
	Coptic	*Coptic*	*Phonetic Extensions Supplement*	*Yi Radicals*	*Letterlike symbols*	*Halfwidth and Fullwidth Form*	*Halfwidth and Fullwidth Form*
	ⲗ	ⲗ	ᶫ	ꓡ	ℒ	ᴸ	L
	2CD0	2CD1	1DAB	A492	2112	FFA4	FF2C
	Numbers Form	*Bopomofo*					
	Ⅼ	ㄥ					
	216C	3125					
M 004D	*Greek and Coptic*	*Greek and Coptic*	*Cyrillic*	*Cyrillic*	*Cherokee*	*Phonetic Extensions*	*Phonetic Extensions*
	M	M	M	ᴍ	M	ᴹ	ᴍ
	03FA	039C	041C	043C	13B7	1D39	1D0D
	Runic	*Letterlike symbols*	*Coptic*	*Numbers Form*	*Halfwidth and Fullwidth Form*	*Coptic*	
	ᛗ	ℳ	ⲙ	M	M	ⲙ	
	16D6	2133	2C99	216F	FF2D	2C98	
N 004E	*Latin Extended-B*	*IPA Extensions*	*Greek and Coptic*	*Phonetic Extensions*	*Halfwidth and Fullwidth Form*	*Coptic*	*Coptic*
	Ɲ	ɴ	N	ᴺ	N	☐	☐
	019D	0274	039D	1D3A	FF2E	2C9A	2C9B
	Phonetic Extensions Supplement	*Letterlike symbols*					
	ᶰ	ℕ					
	1DB0	2115					
O 004F	*Greek and Coptic*	*Cyrillic*	*Armenian*	*Nko*	*Oriya*	*Yi Syllables*	*Yi Radicals*
	O	O	O	□	0	0	0

	039F	041E	0555	07C0	0B20	A132	A4A8
	Limbu	*Halfwidth and Fullwidth Form*	*Coptic*	*Oriya*			
	□	O	□	0			
	1946	FF2F	2C9E	0B66			
P 0050	*Latin Extended-B*	*Greek and Coptic*	*Cyrillic*	*Cherokee*	*Phonetic Extensions*	*Phonetic Extensions*	*Phonetic Extensions*
	Ƥ	Ｐ	Р	Ꮲ	ᵖ	ᴾ	ᵖ
	01A4	03A1	0420	13E2	1D18	1D29	1D3E
	Halfwidth and Fullwidth Form	*Coptic*	*Letterlike symbols*				
	P	□	ℙ				
	FF30	2CA2	2119				
Q 0051	*Cyrillic Supplement*	*Halfwidth and Fullwidth Form*	*Tifinagh*	*Letterlike symbols*			
	ԛ	Ｑ	ⵕ	ℚ			
	051A	FF31	2D55	211A			
R 0052	*Latin Extended-B*	*Latin Extended-B*	*IPA Extensions*	*Cherokee*	*Cherokee*	*Unified Canadian Aboriginal Syllabics*	*Phonetic Extensions*
	Ʀ	Ｒ	ʀ	Ꭱ	Ꭱ	Ꮢ	ᴿ
	01A6	024C	0280	13A1	13D2	1587	1D3F
	Letterlike symbols	*Halfwidth and Fullwidth Form*	*Letterlike symbols*				
	ℛ	R	ℝ				
	211B	FF32	211D				
S 0053	*Cyrillic*	*Armenian*	*Georgian*	*Cherokee*	*Yi Syllables*	*Yi Radicals*	*Vai*
	Ѕ	Ｓ	Ⴝ	Ꮪ	ꑭ	꒪	ⵙ
	0405	054F	10BD	13DA	A1D9	A49A	A576
	Halfwidth and Fullwidth Form	*Malayalam*					
	Ｓ	ട					
	FF33	0D1F					
T 0054	*Latin Extended-B*	*Greek and Coptic*	*Greek and Coptic*	*Greek and Coptic*	*Cyrillic*	*Cyrillic*	*Nko*
	Ƭ	Ｔ	□	□	Т	т	ⵜ
	01AC	03A4	0372	0373	0422	0442	07E0
	Hangul Compatibility Jamo	*Yi Radicals*	*Vai*	*Halfwidth and Fullwidth Form*	*Ethiopic*	*Coptic*	*Coptic*
	ㅜ	꒐	ⵜ	Ｔ	ᴛ	□	□
	315C	A4C4	A50B	FF34	1350	2CA6	2CA7

	Cherokee	*Phonetic Extensions*	*Phonetic Extensions*	*Bopomofo*			
	T	ᵀ	ᴛ	ㄒ			
	13A2	1D40	1D1B	3112			
U 0055	*Armenian*	*Unified Canadian Aboriginal Syllabics*	*Phonetic Extensions*	*Phonetic Extensions*	*Yi Radicals*	*Halfwidth and Fullwidth Form*	
	U	ᑌ	ᵁ	ᵤ	⋃	U	
	054D	144C	1D41	1D1C	A4A4	FF35	
V 0056	*Cherokee*	*Numbers Form*	*Halfwidth and Fullwidth Form*	*Unified Canadian Aboriginal Syllabics*			
	Ꮩ	Ⅴ	V	ᐯ			
	13E4	2164	FF36	142F			
W 0057	*Cyrillic*	*Cyrillic Supplement*	*Cherokee*	*Cherokee*	*Halfwidth and Fullwidth Form*		
	Ꮃ	□	Ꮤ	Ꮃ	W		
	0460	051C	13B3	13D4	FF37		
X 0058	*Greek and Coptic*	*Cyrillic*	*Numbers Form*	*Yi Radicals*	*Halfwidth and Fullwidth Form*	*Coptic*	*Coptic*
	X	Х	Ⅹ	ᚷ	X	□	□
	03A7	0425	2169	A4BC	FF38	2CAC	2CAD
	Runic						
	ᚷ						
	16B7						
Y 0059	*IPA Extensions*	*Greek and Coptic*	*Cyrillic*	*Cyrillic*	*Halfwidth and Fullwidth Form*		
	ʏ	Y	Ⅿ	ӌ	Y		
	028F	03A5	04AE	04AF	FF39		
Z 005A	*Latin Extended-B*	*Latin Extended-B*	*Greek and Coptic*	*Myanmar*	*Cherokee*	*Halfwidth and Fullwidth Form*	*Coptic*
	Ʒ	ƶ	Ζ	□	Ꮓ	Z	□
	0224	01B5	0396	1097	13C3	FF3A	2C8C
	Limbu	*Letterlike symbols*					
	□	ℤ					
	1901	2124					
a 0061	*IPA Extensions*	*Cyrillic*	*Tai Le*	*Phonetic Extensions*	*Phonetic Extensions*	*Halfwidth and Fullwidth Form*	
	ɑ	а	ᥲ	ᵃ	ᵅ	a	
	0251	0430	1972	1D43	1D45	FF41	
b 0062	*Latin Extended-B*	*Latin Extended-B*	*Cyrillic*	*Cyrillic*	*Nko*	*Cherokee*	*Unified Canadian*

							Aboriginal Syllabics
	ƀ	ƀ	ƀ	ƀ	□	ƀ	ƀ
	0185	0184	042C	044C	07D5	13CF	15AF
	Phonetic Extensions	*Halfwidth and Fullwidth Form*					
	ᵇ	b					
	1D47	FF42					
c 0063	*Greek and Coptic*	*Myanmar*	*Tai Le*	*Numbers Form*	*Halfwidth and Fullwidth Form*	*Coptic*	*Cyrillic*
	c	□	□	c	c	□	c
	03F2	1004	1974	217D	FF43	2CA5	0441
d 0064	*Cyrillic Supplement*	*Cyrillic Supplement*	*Unified Canadian Aboriginal Syllabics*	*Phonetic Extensions*	*Numbers Form*	*Yi Radicals*	*Yi Radicals*
	ԁ	ԁ	ԁ	ᵈ	d	ԁ	ԁ
	0500	0501	146F	1D48	217E	A4AD	A4AF
	Halfwidth and Fullwidth Form	*Letterlike symbols*					
	d	ⅆ					
	FF44	2146					
e 0065	*Cyrillic*	*Myanmar*	*Tai Le*	*Phonetic Extensions*	*Letterlike symbols*	*Letterlike symbols*	*Halfwidth and Fullwidth Form*
	e	□	□	ᵉ	e	*e*	e
	0435	1054	1971	1D49	212E	212F	FF45
f 0066	*Halfwidth and Fullwidth Form*	*Phonetic Extensions*					
	f	ᶠ					
	FF46	1DA0					
g 0067	*IPA Extensions*	*Phonetic Extensions*	*Letterlike symbols*	*Halfwidth and Fullwidth Form*	*Phonetic Extensions Supplement*		
	g	ᵍ	ℊ	g	ᵍ		
	0261	1D4D	210A	FF47	1DA2		
h 0068	*Spacing Modifier Letters*	*Cyrillic*	*Cyrillic*	*Georgian*	*Cherokee*	*Letterlike symbols*	*Halfwidth and Fullwidth Form*
	ʰ	h	h	h	h	*h*	h
	02B0	04BA	04BB	10B9	13C2	210E	FF48
i 0069	*Cyrillic*	*Cherokee*	*Phonetic Extensions*	*Numbers Form*	*Halfwidth and Fullwidth Form*	*Letterlike symbols*	

	i	i	i	i	i	i	
	0456	13A5	1D62	2170	FF49	2148	
j 006A	*IPA Extensions*	*Spacing Modifier Letters*	*Greek and Coptic*	*Cyrillic*	*Unified Canadian Aboriginal Syllabics*	*Halfwidth and Fullwidth Form*	*Letterlike symbols*
	j	j	j	j	J	j	j
	029D	02B2	03F3	0458	148E	FF4A	2149
k 006B	*Cyrillic*	*Phonetic Extensions*	*Halfwidth and Fullwidth Form*	*Coptic*			
	к	k	k	□			
	043A	1D4F	FF4B	2C95			
l 006C	*Spacing Modifier Letters*	*Cyrillic*	*Telugu*	*Tai Le*	*Numbers Form*	*Tifinagh*	*Vai*
	l	I	l	□	l	□	□
	02E1	04CF	0C79	1963	217C	2D4F	A621
	Halfwidth and Fullwidth Form	*Yi Radicals*					
	l						
	FF4C	A490					
m 006D	*Phonetic Extensions*	*Numbers Form*	*Halfwidth and Fullwidth Form*				
	m	m	m				
	1D50	217F	FF4D				
n 006E	*Tai Le*	*Halfwidth and Fullwidth Form*					
	□	n					
	1952	FF4E					
o 006F	*Greek and Coptic*	*Cyrillic*	*Armenian*	*Lao*	*Tamil*	*Telugu*	*Myanmar*
	o	o	o	o	o	o	
	03BF	043E	0585	0ED0	0BE6	0C66	1040
	Phonetic Extensions	*Halfwidth and Fullwidth Form*	*Ethiopic*	*Coptic*	*Phonetic Extensions*	*Myanmar*	*New Tai Lue*
	o	o	*o*	□	o	□	□
	1D0F	FF4F	12D0	2C9F	1D3C	101D	19D0
p 0070	*Cyrillic*	*Phonetic Extensions*	*Halfwidth and Fullwidth Form*	*Coptic*			
	p	p	p	□			
	0440	1D56	FF50	2CA3			
q	*Cyrillic*	*Halfwidth*					

0071	*Supplement*	*and Fullwidth Form*					
	□	q					
	051B	FF51					
r 0072	*IPA Extensions*	*Spacing Modifier Letters*	*Phonetic Extensions*	*Halfwidth and Fullwidth Form*			
	ɾ	ʳ	ᵣ	r			
	027C	02B3	1D63	FF52			
s 0073	*Spacing Modifier Letters*	*Cyrillic*	*Halfwidth and Fullwidth Form*	*Unified Canadian Aboriginal Syllabics*	*Latin Extended-D*	*Limbu*	
	ˢ	ѕ	ₛ	ᔆ	□	□	
	02E2	0455	FF53	1506	A731	1949	
t 0074	*IPA Extensions*	*Phonetic Extensions*	*Halfwidth and Fullwidth Form*				
	t	ᵗ	t				
	0288	1D57	FF54				
u 0075	*Phonetic Extensions*	*Halfwidth and Fullwidth Form*	*Phonetic Extensions*				
	ᵘ	u	ᵘ				
	1D64	FF55	1D58				
v 0076	*Greek and Coptic*	*Phonetic Extensions*	*Phonetic Extensions*	*Limbu*	*Number Form*	*Tifinagh*	*Halfwidth and Fullwidth Form*
	ν	ᵛ	ᵥ	□	ⅴ	ⵡ	v
	03BD	1D20	1D65	194E	2174	2D38	FF56
w 0077	*Spacing Modifier Letters*	*Cyrillic Supplement*	*Phonetic Extensions*	*Halfwidth and Fullwidth Form*	*Phonetic Extensions*		
	ʷ	□	w	w	ᵂ		
	02B7	051D	1D21	FF57	1D42		
x 0078	*Spacing Modifier Letters*	*Cyrillic*	*Number Form*	*Bopomofo*	*Halfwidth and Fullwidth Form*	*Limbu*	
	ˣ	х	x	ㄨ	x	□	
	02E3	0445	2179	3128	FF58	194A	
y 0079	*Spacing Modifier Letters*	*Cyrillic*	*Cyrillic*	*Georgian*	*Cherokee*	*Bopomofo*	*Halfwidth and Fullwidth Form*
	ʸ	ү	y	ყ	ʏ	ㄚ	y
	02B8	0423	0443	10B8	13A9	311A	FF59
z 007A	*Latin Extended-B*	*Phonetic Extensions*	*Halfwidth and*	*Unified Canadian*	*Coptic*	*Latin Extended-B*	*Phonetic Extensions*

			Fullwidth Form	*Aboriginal Syllabics*			*Supplement*
	z	z	z	z	□	z	z
	0225	1D22	FF5A	1646	2C8D	01B6	1DBB

Table2: Selected English alphabets for hiding process

Symbols	ASCII	Unicode		
	Secret message 00	Secret message 01	Secret message 10	Secret message 11
A	0041	0391	0410	13AA
B	0042	0392	0412	0181
E	0045	0395	0415	13AC
G	0047	050C	13C0	13B6
H	0048	0397	041D	13BB
I	0049	0399	04C0	0406
M	004D	039C	041C	216F
O	004F	039F	041E	0555
P	0050	0420	03A1	01A4
S	0053	0405	054F	13DA
T	0054	0422	03A4	01AC
j	006A	0458	03F3	029D
o	006F	03BF	1D0F	043E

Table 3: Experimental results of proposed method

Experiment #.	Selected Char.# in Cover	Max.# of bits Can be Embedded in Cover	Ratio of Stego-Doc. Size increasing (%)
1	1313	2626	1.11
2	1465	2930	27.6
3	379	758	17.7
4	1602	3204	0.8
5	4091	8182	2.8
6	3681	7362	2.3
7	672	1344	23
8	268	536	4.2
9	6515	13030	31.7
10	902	1804	0.7
			Avrg. 11.1