# A HYBRID TEXT STEGANOGRAPHY APPROACH UTILIZING UNICODE SPACE CHARACTERS AND ZERO-WIDTH CHARACTER

*Muhammad Aman, Aihab Khan, Basheer Ahmad, Saeeda Kouser*

Iqra University Islamabad Campus
e-mails: m.aman19655@gmail.com, aihab@iqraisb.edu.pk,
drbasheer@iqraisb.edu.pk, saeeda.csit@must.edu.pk
Pakistan

**Abstract:** This paper presents a steganographic approach utilizing Unicode space and Zero-Width Characters. The existing techniques are less robust, not sensitive against steg-analysis and attain low hiding capacity. The proposed technique outperforms the limitations in existing approaches. It tenders high hidden capacity by using lose-less compression algorithm and embedding 4 bits per space using any version of MS Word file as a stego carrier. Moreover, robustness is highly improved by adding multi-layers of security and sensitivity has been created with addition of SHA-1 algorithm. The experimental results verify that the proposed scheme has increased the capacity 4 times and creates 4 times smaller stego-text as compared to existing Unispach method. Moreover, the transparency has not been affected which shows that our approach is best suitable for large messages when high security is required.

**Key words:** Covert Communication; Zero-Width Character; Unicode Space Characters; Permutation; SHA-1.

## 1. INTRODUCTION

Information hiding is the concealment of secret message into cover file [1]. Now a days, due to extensive use of internet and other communication mediums, it is necessary to protect secret information from being accessed by intruder over communication mediums [2]. The applications of information hiding are initiate from old Greek times [3]. At that time, crucial messages were sent by foot, horses and by memorizing. They also used tablets and shaved heads to encode secret messages. The ancient Romans used invisible inks [4] for this purpose. With increasing inventions of new technologies and applications, new threats arouse and to prevent these threats new mechanisms were invented [5]. Generally, Information hiding has two major sub disciplines: Cryptography and Steganography [2, 3].

Cryptography means hidden or secret writing that protects the contents of secrete message [6]. Steganography means covered writing or concealed writing and it is one of the major disciplines among the information hiding methods getting more importance day by day [4]. Steganography is given preference over cryptography because of its ambiguous structure that attracts the intruder to sense the existence of secret information [4, 7]. John Wilkins has given preference to steganography over cryptography because of non-suspicious behaviour of steganography [4, 8]. Steganography is a method that keeps the existence of message secret in cover file and creates a covert communication. It camouflages secret message in the cover file in such a way that without a recipient, no one realizes the presence of concealed information [21, 24]. Steganography techniques are widely applied to English texts [24]. These techniques are categorized into various disciplines according to the utilized features of cover text [4, 14].

On the basis of cover mediums used to embed secret information, steganography is classified into image, text, audio and video steganography [9]. Text steganography is difficult because of less redundancy in text files in comparison with other digital mediums [4,10,11]. On the other hand, text is occupying small space in memory, simple in communication and is widely available over internet in digital form in contrast to other mediums [12]. Text steganography has three main aspects: capacity, robustness and transparency. These parameters are referred as strength measuring elements of any text steganography method.

- Capacity is interpreted as the payload carrying capability of a cover text. The number of bits it can embed in embedding phase is termed as its capacity.
- Robustness is the resistance of a stego-object to modifications, destruction or extraction of concealed secret information by an intruder during communication.
- Transparency is the innocuous look of a stego-object to eavesdropper eliminating the impact of suspicious behavior [4, 13]. It must be the stego-text property to avoid the attention of intruder being an ordinary text to keep the secret information secure.

Text based steganography is classified into three types on basis of concealment into cover file: Format based steganography; Linguistic based steganography and Random and Statistical Generation method [4, 14]. Format based methods modify text or change text formats in cover file by inserting spaces, non-displaying characters, style changing, words changing, lines changing, text resizing, and original features changing in cover files [15-17].

Linguistic steganography conceals secret information by modifying linguistic properties of cover text of a natural language. Linguistic steganography is further divided into two types: syntactic methods and semantic methods [18-20]. Syntactic methods camouflage secret message string by identifying proper places for insertion of full stop (.) and comma (,). For embedding of bit 0, it inserts full stop and for bit

1, it inserts comma [4, 20-22] . Semantic methods hide information by replacing words with their synonyms [4, 20-22].

The proposed work presents a novel format-based open-spaces method defining hybrid approach combining Unispach [23] and Zero-Width Characters [21] approaches in a novel way using word document as a cover file. The presented method overcomes the drawbacks of low embedding capacity and low robustness without affecting the imperceptibility. The secrete string is permuted, compressed and then thinnest width Unicode space characters i.e. Thin, Six-Per-Em, Hair and Zero-Width Character (ZWC) are used collectively for embedding secret bits into inter-word and inter-sentence spaces. Whereas, Unicode's Hair, Six-Per-Em, Punctuation and Thin are used in end-of-line and inter-paragraph spaces to encode payload. The Unicode's selected for embedding payload are similar just like a normal space and does not affect the normal behavior of spaces present in stego-text after embedding secret bits into spaces. The transparency of proposed technique is same as of [23] because insertion of an extra character ZWC by proposed method has no weight and did not increase space length. Whereas, the robustness and capacity are enhanced by merging these characters.

## 2. PROBLEM DEFINITION

Format based steganography approaches [19, 21, 23, 25-33] are hiding secret text either by using inter-word, inter-paragraph, word shift, line shift, features coding, end-of-line, inter-sentence and special characters insertion between the white spaces or by combining these methods as a hybrid solution. The disadvantage of these schemes is low embedding capacity because these methods are concealing either 1 or 2 bits only.

Open space approaches [19, 21, 31] are utilizing special characters to embed information in white spaces. The problem with these approaches is lacking in robustness because of absence of additional security layers. If someone senses the applicability of underlying method, he would be able to extract concealed information.

Feature based approach [25] utilizes diacritics to hide secret message. The main problem of this approach is attraction of intruder attention because of changes or modifications applied to the text. These methods are sensitive to any OCR program and in case of retyping of stego-text, the secret information is lost.

Random and statistical generation methods use grammatical rules (Context Free Grammar) of a certain language which generates cover file automatically and hides secret message within cover characters [11, 16]. The random generation of stego-text gives the impact of encryption rather than steganography by attaining the suspicious behavior.

Semantic methods include text substitution conception [1], multi-text substitution [34], synonyms [35] and context-based substitution in order to hide secret information [36] and requires a complete knowledge of the language. The

main disadvantage of these approaches is low embedding capacity because of concealing 1 bit per substitution. Moreover, these approaches are less robust as once the applicability is known intruder can easily extract the hidden information.

This paper is categorized into various sections. Section III includes problem solution IV contains experimental results and discussion section V is about the conclusion and future directions.

## 3. PROBLEM SOLUTION

This section presents proposed scheme and a layout of information embedding model. The embedding model is shown in Fig. 1.

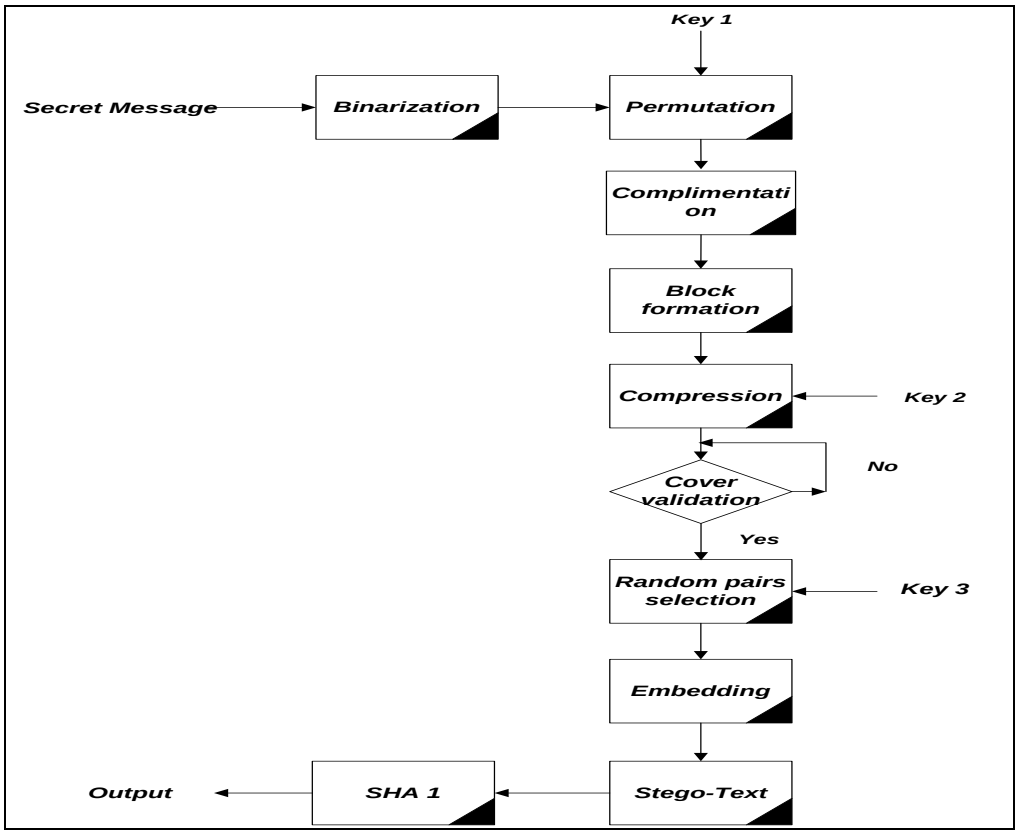


*Fig. 1. Proposed Text Steganography Model*

At first stage, the secret message is converted into a binary string. This binary string is further permuted according to a secret key and then inverted to get a complex binary string to enhance the robustness of secret information. The permutation is accomplished be permuting the secret bits to destroy the character sequences to enhance the complexity level of proposed method. Inversion is also

applied in same context to add another security layer to the secret information before going through embedding process. In inversion process whole bit string is inverted using ones compliment to get a new bit string used for embedding into cover. Furthermore, the algorithm makes blocks of secret string having 4 bits per block to reduce the size of secret bit string by designing a dictionary as shown in Table. 1. The dictionary contains all sixteen possible combinations of 4 bits which are then mapped to 2 bits pairs repeatedly unless whole string is processed. In this way all sixteen combinations are grouped into four mapped combination blocks named as G1, G2, G3 and G4. The group name has been used as a key in embedding and will be used in extraction process to get back the 4 bit block corresponding to each 2 bits extracted group.

*Table 1. Shows compression dictionary*

| ***Combinations*** | ***Mapping*** | ***Group*** |
|---|---|---|
| *0000* | *00* | *G1* |
| *0001* | *01* | *G1* |
| *0010* | *10* | *G1* |
| *0011* | *11* | *G1* |
| *0100* | *00* | *G2* |
| *0101* | *01* | *G2* |
| *0110* | *10* | *G2* |
| *0111* | *11* | *G2* |
| *1000* | *00* | *G3* |
| *1001* | *01* | *G3* |
| *1010* | *10* | *G3* |
| *1011* | *11* | *G3* |
| *1100* | *00* | *G4* |
| *1101* | *01* | *G4* |
| *1110* | *10* | *G4* |
| *1111* | *11* | *G4* |

Each pair of 2 bits is mapped with the four different combinations of 4 bits groups as shown in Table 1. These four groups are constructed depending on the left most two bits pair that remains unchanged for each group. For example G1 is based on the bits pair ‘00’, G2 is based on ‘01’ and so on.

The algorithm scans the secret blocks and reduces each 4 bits block to 2 bits pair according to dictionary in secret string and creates a key with group name. The 2 bits representing group name are excluded from embedding process and only right handed 2 bits pairs are combined to create a new bit string by reducing the size of

secret bits string to half. This mapping and creation of new half-length bit string implements the concept of compression by proposed method. Hence, the proposed method is truly a multi-layered architecture that enhances the capacity, security and robustness significantly.

### 3.1. Embedding Process

The Microsoft word document spaces are classified into two groups: (A) inter-word and inter-sentence spaces (B) end-of-line and inter-paragraph spaces. For group A, the three smallest widths Unicode space characters Thin, Six-Per-Em and Hair are selected and embedded into inter-word and inter-sentence spaces according to randomly selected pair as shown in Fig. 2. This selection is made because these Unicode characters are just like a normal space and do not give the impact of anything suspicious. In addition, at the same time to encode next 2 bits payload, a ZWC is inserted to the same space as it is a width-less character and does not affect the overall space behaviour. For this purpose another random pair of 2bits is selected from secret message string, if the selected pair is '00' then no ZWC is inserted before or after the combination of Unicode and normal space, if the selected pair is '01' then ZWC is inserted after the combination, if the selected pair is '10' then ZWC is inserted before and if '11' then ZWC is inserted before and after the combination. The insertion of ZWC continues in parallel with Unicode characters while embedding in inter-word and inter-sentence spacing.

For group B, the four smallest width Unicode space characters Hair space, Six-Per-Em, Punctuation and thin are randomly selected and embedded in end-of-line and inter-paragraph spaces to encode a 2 bits payload per space in cover file as shown in Fig. 3. The insertion of these character is also dependent on bits pairs randomly selected from secret bits string as mentioned in Fig. 3.

| Combination | Sequence |
|---|---|
| Normal | 00 |
| Thin+Normal | 01 |
| Six-Per-Em+Normal | 10 |
| Hair+Normal | 11 |

*Fig. 2. shows group A.*

| Character | Sequence |
|---|---|
| Hair | 00 |
| Six-Per-Em | 01 |
| Punctuation | 10 |
| Thin | 11 |

*Fig. 3. shows group B.*

### 3.2. Proposed Algorithm

This section is about the embedding and extraction algorithms. Embedding algorithm is used to conceal the secret data to the cover text at sender side. The resulting cover file is transmitted over a communication channel and receiver extracts the secret data by applying extraction algorithm and secret key on stego-object.

### 3.2.1. Embedding Algorithm

**Start:**
**Input:** Sec_ Message, Cover_ Text;
**Procedure:**
1. Convert Sec_ Message to Binary_ String;
2. Permuted_ String = Permute (Binary_ String);
3. Complemented_ String = one's Complement (Permuted_ String);
4. Compressed_ String = Compress (Complemented_ String); as follows:

Repeat:

- Divide complemented_ String to block of 4 each;
- Sub Divide each block to 2 part having 2 bits each;
- Assign group No. to 1st 2 bits according to group formation;
- Generate compressed_ String of 2nd part of completed_ String;

Until (end of Complemented_ String. length)

5. Valid_ Cover_ Text (Cover_ text):
   - IF (No_ of_ Spaces (Cover_ Text) >= ¼(Compressed_ String. LENGTH)
     Return TRUE;
   - ELSE
     Return FALSE
6. Embedding (Cover_ Text, Compressed_ String) //For inter-word and inter-sentence spaces
   - Generate Random_ Series_ Numbers;
   - FOR (each 2 bits pair of Compressed_ String)
     - IF(space= 'inter-word/inter-sentence')
     - Embed Unicode_ Char into white space of cover-text w.r.t 2 bit pair value according to Random_ Series_ Numbers;
   - FOR (each 2 bits pair of Compressed_ String)
     - Embed ZWC to the left, right or both sides of Unicode w.r.t 2 bit pair value randomly;
   - Else // For inter-sentence and inter-paragraph spaces
     - Embed Unicode_Character in white space of cover-text w.r.t 2 bit pair value randomly;
   - Update cover_ Text after Embedding;
7. Calculate hash Function of Cover_ Text;

**Output** Stego_ Text, Keys, hash- value;

**End:**

### 3.2.2. Extraction Algorithm

**Start:**
**Input:** Stego_ Text, Keys, hash_value;
**Procedure:**
1. Calculate Hash function value to check authenticity of Stego_ Text;
2. Extract secret bits from stego_ Text using key;
3. Decompress 2 secret bits to 4 by combining 2 extracted bits pair with 2 bits pair obtained using groups dictionary;
4. Generate decompressed bit string;
5. Find one's complement of decompressed String;
6. De-permute the complemented String;
7. Get the byte String of de-permuted Binary_ string
8. The resultant string is extracted Secret_ Message

**Output:** Secret_ Message
**End**

The proposed method is implemented having two phases; Embedding and extraction phase as described in algorithms. The system first embeds the secret bits by applying all the predefined functions to a MS word file. Furthermore, it extracts the secret message by applying a reverse procedure to the stego file. The screen shot of implemented system is provided in figure 4.

## 4. RESULTS AND DISCUSSIONS

This section contains the experiments results carried out for testing of proposed system and results analysis. The proposed algorithm is tested against two basic parameters of text steganography: capacity and robustness. The detailed analysis is illustrated below to have a close look on the achieved results that overcome the limitations exists in text steganography techniques.

### 4.1. Capacity Analysis

Capacity of the proposed system is measured in terms of the ratio obtained as a result of hidden bits amount per space in cover file. The capacity results are compared with existing work [23] which showed that the proposed method has higher embedding capacity. For experiments we have taken 11different sets of payloads and embedded those to cover-text with minimum number of spaces count using both methods which showed that the proposed method requires 4 times lesser spaces for a fix size of payload than Unispach.

It is observed that Unispach embeds 2 bits / space and the proposed technique embeds 4 bits / space but, the proposed method also compresses the string to half before embedding. So, mathematically speaking if we overall calculate the number of bits per space, the proposed technique embeds 8 bits / space that is 4 times greater

than the Unispach. It shows that the proposed method is rich enough than existing methods in terms of capacity analysis. Fig.5 shows the results of our approach in comparison with Unispach [23] in terms of space count of both techniques verses fixed size payload.

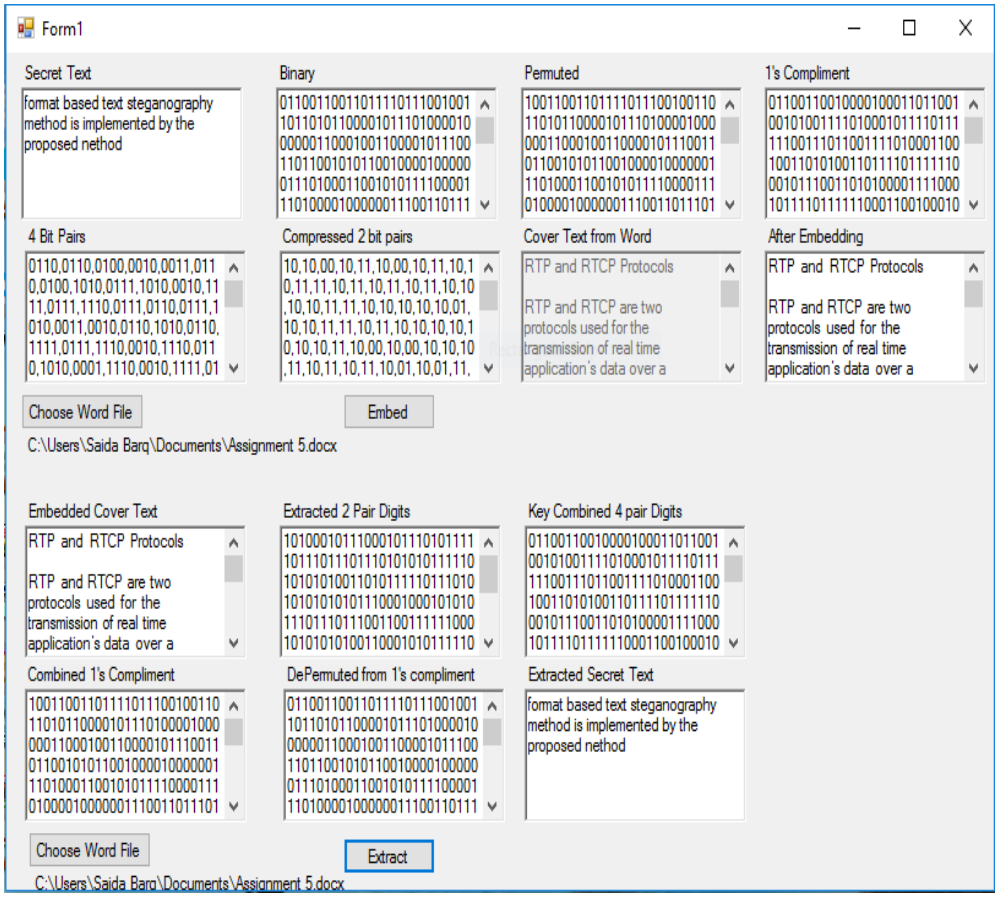


*Fig. 4. Screen shot of implemented Text Steganography System.*

### 4.2. Stego-text Size Analysis

The experimental results for maximum steg-text size analysis show that for a fixed size of secret message, the average increment in output stego-text size is 4 times smaller than Unispach followed by White-steg, SNOW, Spacemimic and WbStego4open respectively. Fig. 6 depicts that all these approaches required larger cover files for embedding secret data as they required more number of white spaces as compared to proposed technique. It is so because the proposed technique embeds more number of secret bits per space as compared to existing techniques.

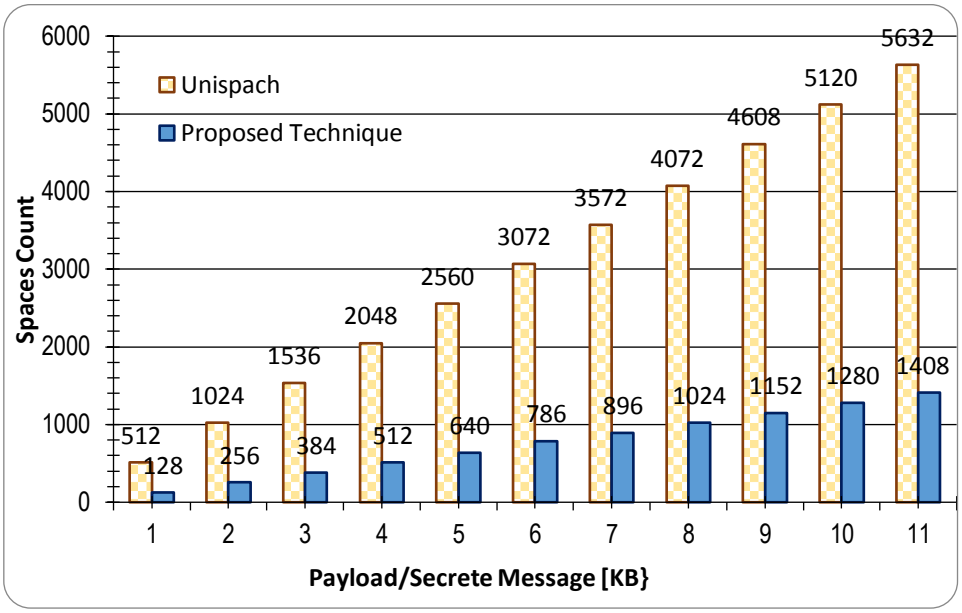


*Fig. 5. Capacity comparison of Unispach versus Proposed Technique*

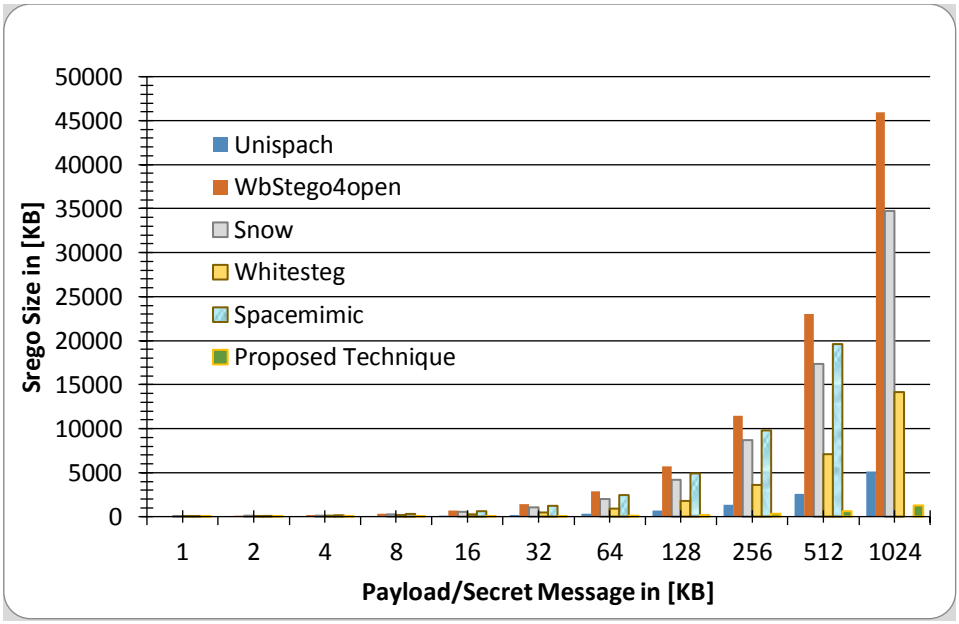


*Fig. 6. Stego-text size comparison for a fix size of payload.*

Table. 2 further clarifies the results of average stego-text sizes created as output for different techniques used for a fixed size of payload by embedding in minimum size of cover text as shown in Fig. 6.

*Table 2. Stego-text size against different size of payload*

| ***S.M [KB]*** | ***Unispach*** | ***WbStego-4open*** | ***Snow*** | ***Spacemi-mic*** | ***Whitesteg*** | ***ProposedTechnique*** |
|---|---|---|---|---|---|---|
| *1* | *9* | *47* | *45* | *40* | *25* | *3* |
| *2* | *16* | *92* | *80* | *80* | *43* | *4* |
| *4* | *29* | *183* | *134* | *154* | *81* | *7* |
| *8* | *56* | *362* | *266* | *307* | *146* | *14* |
| *16* | *103* | *772* | *527* | *616* | *274* | *26* |
| *32* | *187* | *1437* | *1020* | *1227* | *500* | *47* |
| *64* | *348* | *2871* | *1998* | *1446* | *930* | *87* |
| *128* | *667* | *5743* | *4150* | *4909* | *1810* | *167* |
| *256* | *1320* | *11493* | *8709* | *9797* | *3608* | *330* |
| *512* | *2595* | *22992* | *17383* | *19607* | *7105* | *649* |
| *1024* | *5125* | *45979* | *34731* | * | *14127* | *1282* |

* spacemimic method fails to process S.M of 1024 KB and above.

### 4.3. Robustness Analysis

Robustness is the resistive power of an algorithm to various steg-analysis attacks to access crucial information without having secret key and permission. The information must only be extracted by authorized persons having secret key. For this purpose, the output stego-text produced is tested against various types of steg-analysis attacks which reveal the strength of proposed approach against the intruder illegal intervention and actions. The experiments show that output stego-text produced is highly robust in comparison with Unispach, if it is tried to temper. It intimates the receiver about the illegal action, if performed by any unauthorized personnel, during communication.

***Visual attack***

The visual attack is worthless because the stego-text looks normal and human eye cannot detect any irregular pattern.

***Statistical attack***

The proposed approach is robust to statistical attacks because the behaviour of word document is normal. It would not be possible to detect the secret message by analysis of consecutive neighbour spaces difference by the intruder as there is no ambiguity among spaces created by the proposed method. Suppose if someone is successful to read the characters in spaces, it is still impossible to extract secret

information because the secret string in permuted, complemented, compressed and bit pairs are selected randomly for embedding with different secret keys making it highly secure and robust.

***Format, insertion, deletion, reordering, re-composition and replacement attack***

The stego-text is capable to address all these attacks and will pop up a message about illegal action of intruder to ensure the integrity of stego and will inform the receiver. In case of any illegal action, it displays a message shown in Fig. 7 to the receiver and demands for resending of secret information by the sender.

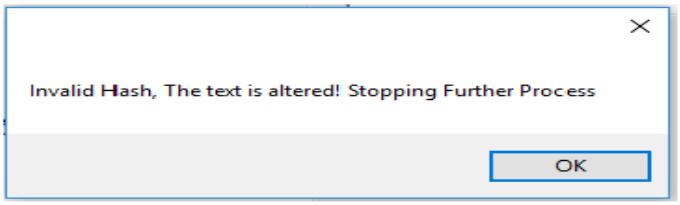


*Fig. 7. Pop up Message for any above attack*

### 4.4. Transparency / imperceptibility

Stego-text produced as output is highly Imperceptible and no suspiciousness can be detected with human eye. Only authorized person can access the information and capable to extract it successfully. The experimental results have proved that the stego-text created has same imperceptibility as compare to Unispach. A secret message has been embedded through both techniques using same cover-text and resultant stego-text created of Proposed technique as shown in Fig. 8 is completely identical to Unispach as shown in Fig. 9.

China in 190 A.D. was nearing chaos. The Second Han Dynasty, which had ruled for the previous 165 years, was dying. Law throughout the country was not coming from the Emperor but from generals who grabbed paver where and when they could. At first no single general controlled more than a handful of states at most and many states were controlled by no one at all. As time went on a few generals managed to gradually expand their rule until by 215 A.D. China was divided into three kingdoms called Wei, Wu and Shu. The ruler of each desperately trying to consolidate the entire country under himself. This era is referred to as the Three Kingdoms Period. Romance of The Three Kingdoms is a simulation game that traces China from the chaos with which the Three Kingdoms began to its rule by one general. The player takes control of a master, a general capable of commanding as many states as he can acquire, and, if successful, unifies Chine. As many as eight may play, but only one can succeed. There are five chronologically arranged scenarios. The first has China in its most disorganized period and the last has virtually all of China controlled by one of three generals. The precise requirements for success in each of these scenarios differs but in all cases the goal is to rule as many states as possible. After the completion of any scenario but number five the game will automatically advance to the next. You may start the game at any scenario. Koei's Romance of The Three Kingdoms is based on a historical novel of the same name written in the Fourteenth Century, which was in turn based on a more serious official work of history by Ch'en Shou who chronicled major historical events in China from 220 to 265 A.D. Your master strives to unite China. You must enlist the help of others, fight well and negotiate shrewdly.

*Fig. 8. Stego-text created by proposed technique*

China in 190 A.D. was nearing chaos. The Second Han Dynasty, which had ruled for the previous 165 years, was dying. Law throughout the country was not coming from the Emperor but from generals who grabbed paver where and when they could. At first no single general controlled more than a handful of states at most and many states were controlled by no one at all. As time went on a few generals managed to gradually expand their rule until by 215 A.D. China was divided into three kingdoms called Wei, Wu and Shu. The ruler of each desperately trying to consolidate the entire country under himself. This era is referred to as the Three Kingdoms Period. Romance of The Three Kingdoms is a simulation game that traces China from the chaos with which the Three Kingdoms began to its rule by one general. The player takes control of a master, a general capable of commanding as many states as he can acquire, and, if successful, unifies Chine. As many as eight may play, but only one can succeed. There are five chronologically arranged scenarios. The first has China in its most disorganized period and the last has virtually all of China controlled by one of three generals. The precise requirements for success in each of these scenarios differs but in all cases the goal is to rule as many states as possible. After the completion of any scenario but number five the game will automatically advance to the next. You may start the game at any scenario. Koei's Romance of The Three Kingdoms is based on a historical novel of the same name written in the Fourteenth Century, which was in turn based on a more serious official work of history by Ch'en Shou who chronicled major historical events in China from 220 to 265 A.D. Your master strives to unite China. You must enlist the help of others, fight well and negotiate shrewdly.

*Fig. 9. Stego-text created by Unispach*

The strings similarity of different words has been checked through jaro-winkler score which is same in comparison to Unispach while extra character (ZWC) addition in spaces does not affect the visual appearance as shown in Fig. 10. The Jaro-Winkler score of '1' depicts that the cover and stego-texts are exactly same and if the value is '0', it shows that the strings are totally different. The proposed method attains the score of '1' showing that it does not compromise on the transparency of stego-text.

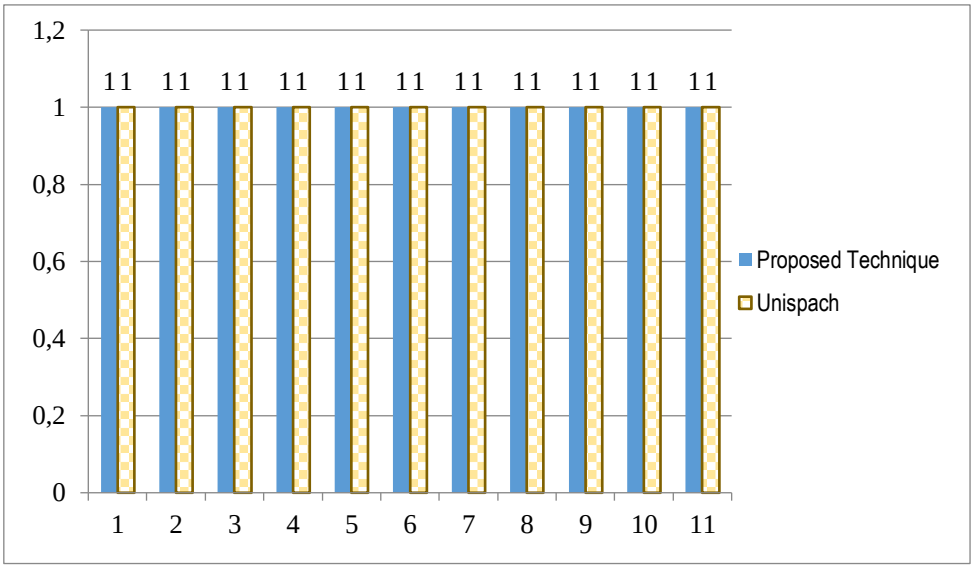


*Fig. 10. Transparency analysis of Unispach versus Proposed Technique*

## 5. CONCLUSION AND FUTURE WORK

The proposed novel approach is open-space format-based text steganography method that utilizes the white spaces to embed the secret information. MS word document is utilized as a cover object that carries the secret information on communication channel. Moreover, all the versions of MS word are compatible for proposed method to use as a cover file. The proposed method is robust, transparent and capable of attaining higher capacity as compared to Unispach and other existing text steganography techniques. Moreover, the experimental analysis depicts that it is robust, secure and perceptible. The results show that the concealing capacity is enhanced from 2bits/space to 4 bits/ space i.e. it is doubled after reduction of string to half through self-designed lose-less compression technique.

In future, the capacity and other parameters like imperceptibility, robustness and security could be enhanced by adding some novel features to the work. In this regard, an encryption algorithm can be a good addition to add another standard security layer to the secret text.

## REFERENCES

[1] S. Bo, D. Xiaoyun, L. Gongshen, Z. Hao, and X. Jing, "An Information Hiding method for Text by Substituted Conception," in *Information Science and Engineering (ISISE), 2012 International Symposium on*, 2012, pp. 131-135.

[2] R. Gupta, "Information Hiding and Attacks: Review," *arXiv preprint arXiv:1404.4141*, 2014.

[3] Z. Jalil, "Copyright protection of plain text using digital watermarking," National University of Computer and Emerging Sciences Islamabad, 2010.

[4] S. M. Thampi, "Information hiding techniques: A tutorial review," *arXiv preprint arXiv:0802.3746*, 2008.

[5] L. Y. Por, T. Ang, and B. Delina, "Whitesteg: a new scheme in information hiding using text steganography," *WSEAS Transactions on Computers*, vol. 7, pp. 735-745, 2008.

[6] "“https://en.wikipedia.org/wiki/Cryptography”, retrieved 09 March, 2016".

[7] V. Ganeshkumar and R. L. Koggalage, "A language independent algorithm to send secret messages using steganography," in *Advances in ICT for Emerging Regions (ICTer), 2010 International Conference on*, 2010, pp. 15-21.

[8] F. A. Petitcolas, R. J. Anderson, and M. G. Kuhn, "Information hiding-a survey," *Proceedings of the IEEE*, vol. 87, pp. 1062-1078, 1999.

[9] J.-C. Chuang and Y.-C. Hu, "Data Hiding on Text Using Big-5 Code," *International Journal of Security and Its Applications*, vol. 6, pp. 173-178, 2012.

[10] U. Khadim, A. Khan, B. Ahmad, and A. Khan, "Information Hiding in Text to Improve Performance for Word Document," *International Journal of Technology and Research*, vol. 3, p. 50, 2015.

[11] I. Banerjee, S. Bhattacharyya, and G. Sanyal, "Novel text steganography through special code generation," in *Proceedings of International Conference on Systemics, Cybernetics and Informatics (ICSCI-2011), Hyderabad, India*, 2011.

[12] A. Al-Azawi and M. A. Fadhil, "Arabic text steganography using kashida extensions with huffmancode," *Applied Sci*, vol. 10, pp. 436-439, 2010.

[13] A. Gutub and M. Fattani, "A novel Arabic text steganography method using letter points and extensions," *World Academy of Science, Engineering and Technology*, vol. 27, pp. 28-31, 2007.

[14] D. Huang and H. Yan, "Interword distance changes represented by sine waves for watermarking text images," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 11, pp. 1237-1245, 2001.

[15] Y. M. Alginahi, M. N. Kabir, and O. Tayan, "An enhanced Kashida-based watermarking approach for Arabic text-documents," in *Electronics, Computer and Computation (ICECCO), 2013 International Conference on*, 2013, pp. 301-304.

[16] S. Kingslin and N. Kavitha, "Evaluative Approach towards Text Steganographic Techniques," *Indian Journal of Science and Technology*, vol. 8, 2015.

[17] R. Saniei and K. Faez, "The Security of Arithmetic Compression Based Text Steganography Method," *International Journal of Electrical and Computer Engineering*, vol. 3, p. 797, 2013.

[18] R. S. R. Prasad and K. Alla, "A new approach to Telugu text steganography," in *2011 IEEE Symposium on Wireless Technology and Applications (ISWTA)*, 2011, pp. 60-65.

[19] L. Por, B. Delina, Q. Li, S. Chen, and A. Xu, "Information hiding: A new approach in text steganography," in *WSEAS International Conference. Proceedings. Mathematics and Computers in Science and Engineering*, 2008.

[20] K. Bennett, "Linguistic steganography: Survey, analysis, and robustness concerns for hiding information in text," 2004.

[21] A. Odeh and K. M. Elleithy, "Steganography in Text by Merge ZWC and Space Character," 2012.

[22] D. Zhang and H. Zhong, "A text hiding method using multiple-base notational system with high embedding capacity," in *Image and Signal Processing (CISP), 2014 7th International Congress on*, 2014, pp. 622-627.

[23] L. Y. Por, K. Wong, and K. O. Chee, "UniSpaCh: A text-based data hiding method using Unicode space characters," *Journal of Systems and Software*, vol. 85, pp. 1075-1082, 2012.

[24] A. T. Abbasi, S. N. Naqvi, A. Khan, and B. Ahmad, "Urdu text steganography: Utilizing isolated letters," 2015.

[25] I. Stojanov, A. Mileva, and I. Stojanovic, "A New Property Coding in Text Steganography of Microsoft Word Documents," 2014.

[26] S. Roy and M. Manasmita, "A novel approach to format based text steganography," in *Proceedings of the 2011 International Conference on Communication, Computing & Security*, 2011, pp. 511-516.

[27] M. Shirali-Shahreza, "Pseudo-space Persian/Arabic text steganography," in *Computers and Communications, 2008. ISCC 2008. IEEE Symposium on*, 2008, pp. 864-868.

[28] M. P. Uddin, M. Saha, S. J. Ferdousi, M. I. Afjal, and M. A. Marjan, "Developing an efficient solution to information hiding through text steganography along with cryptography," in *Strategic Technology (IFOST), 2014 9th International Forum on*, 2014, pp. 14-17.

[29] I. Kamel and S. Banawan, "Hiding information in the placement of maneuverable words," in *Innovations in Information Technology (IIT), 2012 International Conference on*, 2012, pp. 255-260.

[30] M. Shirali-Shahreza and S. Shirali-Shahreza, "Steganography in TeX documents," in *Intelligent System and Knowledge Engineering, 2008. ISKE 2008. 3rd International Conference on*, 2008, pp. 1363-1366.

[31] A. E. Ali, "A new text steganography method by using non-printing unicode characters," *Eng. & Tech. Journal*, vol. 28, 2010.

[32] M. L. Bensaad and M. B. Yagoubi, "High capacity diacritics-based method for information hiding in Arabic text," in *Innovations in Information Technology (IIT), 2011 International Conference on*, 2011, pp. 433-436.

[33] A. A.-A. Gutub, L. Ghouti, A. A. Amin, T. M. Alkharobi, and M. K. Ibrahim, "Utilizing Extension Character'Kashida'with Pointed Letters for Arabic Text Digital Watermarking," in *SECRYPT*, 2007, pp. 329-332.

[34] Y. Shu, L. Liu, W. Tian, and X. Miao, "Algorithm for information hiding in optional multi-text," *Procedia Engineering*, vol. 15, pp. 3936-3941, 2011.

[35] M. H. Shirali-Shahreza and M. Shirali-Shahreza, "A new synonym text steganography," in *2008 International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, 2008.

[36] F. Wang, L. Huang, Z. Chen, W. Yang, and H. Miao, "A novel text steganography by context-based equivalent substitution," in *Signal Processing, Communication and Computing (ICSPCC), 2013 IEEE International Conference on*, 2013, pp. 1-6.

### *Information about the authors:*

**Muhammad Aman** – MSCS Student at Iqra University Islamabad Campus implemented this project for his final year thesis. The area of interest is information security.

**Dr. Aihab Khan** – working as Associate Professor at Iqra University Islamabad Campus. The project has been implemented under his supervision. His area of interest is information security, watermarking and software quality assurance.

**Dr. Basheer Ahmad** – working as Professor of statistics and HOD of Management Sciences Department at Iqra University Islamabad Campus.

**Saeeda Kouser** – MSCS student at Iqra University Islamabad Campus and working as lecturer at Mirpur University of Science and Technology (MUST), Mirpur AJK.

**Manuscript received on 15 January 2017**

**Revised manuscript received on 24 February 2017**