Czech Text Steganography Method by Selective Hiding Technique

Sungkrityayan Khan, Rishi Sankineni, Padmavathi Balagurunathan, Nujerlla Suresh Divya Shree, Abinaya Balasubramanian, *Member, IAENG*

***Abstract*— In this paper we have presented a novel approach of steganography which is suitable for Czech texts. The approach can be assorted under selective hiding method of text steganography. This approach involves the concealing of the information bits clandestinely within the letters gaining from their inherited points. To ascertain the specific letters holding the secluded bits, the propounded technique considers the two features, the redundant Czech extension characters and the extensions of the points in the letters. The un-pointed letters with extension are used by us to hold the secluded bit 'zero' and the pointed letters having an extension to hold the 'one' bit. This specific steganography technique is found similar to other languages having alike text to Czech such as Esperanto, Vietnamese etc.**

***Index Terms*— Czech Text, Text Steganography, Information Security, Cryptography, Text watermarking, Feature coding**

I. Introduction

In the present advancement of the electronic area, steganography plays a very essential role in concealing the data, by hiding it in the extra bits of any unremarkable media[2]. Its principle concern is to keep the message protected in a concealed way by preserving the integrity of the cover media. There are a lot of redundant bits in image, sound and text files which are unused, are replaced by steganography. It's an entirely different methodology from the encryption technique. Encryption furnishes the message opaque to the outsiders by various text transformations, whereas steganography conceals the very presence of data[12].

Security, robustness and capacity[1], mainly affects steganography and its utilities. Capacity indicates the amount of information that can be contained in the form of data bits in the cover media. Security associates to the power of an intruder to estimate the hidden information easily. The characteristics of destroying or modifying the unseen data associates it to robustness.

Though cryptography and watermarking have the same overlapping usages like steganography, yet steganography is quite different from them[12]. The security aspect of steganography mainly relates to the mechanism of concealing the noesis about the data being present in the cover media, whereas on the other hand, cryptography relates to data encryption in the form of cipher-text and decodes it without any sort of authorization; i.e., cryptography concentrates on challenging the decryption process while steganography appends the feature of finding whether there is any concealed data or not. On the other hand, watermarking isn't similar like steganography, in the sense that it aims in protecting the cover media from any alterations with no true stress on secrecy. It can be noticed that steganography is highly focusing on robustness and nearly no security.

Steganography has got numerous usages in different kinds of applications. Medical doctors implements steganography to conceal patient data within various kinds of medical images. In case corruption occurs due to bad connection or transmission, it can engraft the correct data or audio files. Army even implements it to establish a secure channel for private communication, nevertheless, it doesn't ensure the fact that the communication was encountered or the data bits are concealed. This particular feature makes steganography as an entirely new way of encryption or cryptography[6].

Videos[3], Pictures[7], and often sounds[5] are implemented as the cover media in the field of steganography. Although, text steganography is seldomly implemented as because detection of the redundant bits present in the text file, is very hard[3,10]. Text documents structural representation is comparatively similar to what we view, whereas in all other types of cover media, the structural representation is entirely different than what we actually

Manuscript received March 16, 2015; revised April 13, 2015.

Sungkrityayan Khan, is a student of SRM University, Chennai, Tamil Nadu, India. He is a member of the IAENG, IACSIT and life time member of IAETSD and (phone: +919940455822; e-mail: sungkrityayankhan@ gmail.com).

Rishi Sankineni, is a student of SRM University,Chennai, Tamil Nadu, India. He is a student of the Department of Computer Science, SRM University. He is a member of IAENG.(phone: +918148814092; e-mail: rishi.sankineni@gmail.com).

Padmavathi Balagurunathan, is the Head Of The Department Of Computer Science and Engineering, SRM University, City Campus, Chennai, Tamil Nadu, India. She is a member of the IAENG, MISTE, MCSI (phone: +918754487492; e-mail: padmas9169@yahoo.co.in).

Nujerlla Suresh Divya Shree is a student of SRM University,Chennai, Tamil Nadu, India. She is a student of the Department of Computer Science, SRM University. She is a member of the IAENG (phone: +918939818293; e-mail: nsdivya_94@yahoo.com).

Abinaya Balasubramanian, is a student of SRM University,Chennai, Tamil Nadu, India. She is a student of the Department of Computer Science, SRM University. She is a member of the IAENG (phone:+919500196704; e-mail: abinaya14_smile@yahoo.in).

ISBN: 978-988-19253-4-3
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

view. The main advantage of favoring text steganography over all other cover medias is that it occupies lesser memory and it has got a very simple communication[4].

A very important and a significant role is played by the languages and their structures in the favored steganographic system. Commonly single technique is not used for all languages[12]. Numerous techniques used for concealing the information within the electronic Czech text files have been explained by us. An existing method[4] for hiding information in Arabic text is given in the paper. We have discussed our new Czech text steganography method using selective hiding technique. At the last section, we have given the conclusion.

II. Text Steganography Methods

Numerous researches solely related to concealing the information in texts have been implemented. Numerous techniques in the form of examples are exhibited in this section below:-

A. Specific Characters Present in Words

Concealing of information can be done by selecting characters in certain words. The approach mentioned above can vary from being very simple to extremely complicated solely depending on the description given. Let us say, for example, in the simplest form, the first words from every paragraph are selected in such a way, that by placing the initial characters of the picked out words side by side, the concealed information can be extracted by us [4]. Furthermore, elaborated instance can be given by choosing the first letter from the first word, then second letter from the second word and consecutively it goes on, to conceal the data in the cover text.

B. Documents Based On HTML

To conceal the secret information, HTML Tags [8] can also be used, since they exhibit case insensitivity. For example, the tags <p align="center">, <p aLigN="center">, <p align="cenTER">, and <p align="Center"> are all likewise valid. In HTML documents, steganography can be performed by altering the lowercase or upper case letters in the tags of the document. Extraction of data can be by comparing these tags words with words in general case. The security of this HTML steganography can be enhanced by selecting a sequence function of a certain letter. For example, the third upper case letter present within the tags, where most tags should have several randomly altered letters so that the eavesdropper gets confused.

C. Line Shifting and Word Shifting

Shifting of text lines and words vertically and horizontally [11] respectively, may help in concealing some information. The security of this line and word shifting technique solely rests on how the distances between words and lines are varied to confuse the intruders. In this method the lines are shifted up or down somewhat with a fixed space (let us assume 0.002 inch) and alters the distances between words, as per the specified concealed information. This text shifting steganography counts on building visual shapes for data to be hidden in spaces. The technique is suitable for printed texts as in newspapers, since it encounters problems against robustness. Whenever, the text is electronically altered or rewritten, there is a great chance for the hidden/concealed information to be destroyed. Moreover, when we are using programs based on character recognition, for example OCR, the visual shapes that are hiding information cannot be traced accurately or they gets lost.

D. Spaces and Abbreviations

By using, abbreviations and spaces [10], text steganography can conceal very small information in the text. Say for example, “a very less number of bits can be hidden in a file size of various kilobytes” [4]. In Space Steganography method, additional white-spaces between the words, or paragraph of the text or at the end of lines are added to hide the data [10]. This approach is very flexible and can be used with any text. It does not reveal the secret information to the normal reader, its security is good.

Nevertheless, its capacity and robustness is comparatively low. The above approach cannot hide a lot of information and few electronic text editors automatically removes the additional whitespaces.

E. Methods Related To Character Features

For addressing the problems of the OCR usage of the previous shifting approach, character feature methods [9] are advised. Very few characteristics of the text characters are changed in the Character feature steganography. For instance, the most important bits of a few characters are covered to hold the hidden data bits [9]. Character feature methods can hold a bulk amount of secret information, even the normal readers have got no knowledge about the presence of such hidden information in the text.

III. Related Work

An unique security method solely related to character feature for Arabic and Persian letters was propounded by Shirali- Shahreza [4]. Their propounded approach is mainly based on the points present in the Arabic and Persian letters [13], which are somewhat nearly alike. The principal concern in this study will be on the steganographic technique concerned to Arabic language. The amount of pointed letters differ widely in Arabic and English language. Points exists in only two letters in English, i.e., lowercase "i" and lowercase "j", while 15 letters in Arabic script possesses points , which is given in Fig. 1. This large number of presence of points in Arabic text have made it remarkably noticeable in any Arabic text and can be applied for steganography and information security as suggested by Shirali-Shahreza in their “new approach to Persian/Arabic text steganography” [4]

ISBN: 978-988-19253-4-3
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

Unpointed Letters	Pointed Letters
ر د ح ا	ث ت ب
ص س	ز ذ خ ج
ك ع ط	ض ش
ل و ه م	ف غ ظ

Fig 1 Arabic Letters

Shirali and Shahreza [4] has suggested a way of concealing the information bits in the points of the letter. To be very certain, they have used the points present in the pointed letter to conceal the information bits. Initially, the concealed information is considered in a binary way with the first several bits (assume, 15 bits) to show how long the hidden data bits are, that are to be stored. Secondly, text that is used as a cover media, is scanned thoroughly. The location of the point in the pointed letters may be affected by the hidden data bits. The point slightly slips up, if the hidden data bit's value is found to be 1; otherwise, the point of the cover text location remains the same. Below in Fig. 2, the point shifting phenomenon for the Arabic letter 'Fa' is shown. "As a result to confuse the readers, after all the data bits are concealed, the points of the other letters in the text are also changed haphazardly"[4]. As we have mentioned previously, the size of concealed data bits is known to us and also hidden in the first 15 bits. This method of hiding the data has got a lot of advantages including security and capacity; it can store a bulk amount of data bits very securely within any Arabic text. However, it has got drawbacks in robustness, thus making it very impractical for use. For example, the hidden data bits are lost during the process of scanning etc. Due to the use of only one font, the output text has got a fixed frame. Hence it would be far more appropriate to name this process as Watermarking.

Fig 2 Shifting-up of the point of Arabic letter 'Fa'

IV. Proposed Methodology

In this section, an entirely new method of ensconcing the data using any letter instead of only using the pointed ones only has been inferred by us. The secret bit 'zero' is stored in the un-pointed letters and the secret information bit 'one' is stored in the pointed letters. The letter extensions do not affect the inscription content. The standard hexadecimal code for it is 0640 in the Unicode System. To the matter of fact, this extension character of the Czech script in electronic inscription is surfeit and is only advised to be applied for formatting and arrangement purposes.

The only unison that is to be followed while making the use of extension is that, extension characters cannot be used with all letters because of their position in words and Czech inscription style. Between the inter-links of the Czech text, the extensions can be appended; i.e. in front of the letters or after the conclusion of the words, extensions can't be encompassed. Our propounded approach presumes that, at any time if there is no prospect of an extension to be found in the text or if it is found deliberately without an extension it is accounted not to carry any secluded information in the form of bits.

Our propounded technique gives the user an option of appending the extensions anteriorly or after the letters. To assure that the methodology remains unvarying all throughout the entire approach, it is important to keep the extension position same.

V. Results

Let us assume, that we have appended the extension after the letters. We have given an example below in Fig. 3, for the better comprehension of the propounded approach. Primarily the secrets bits that are to be concealed in the cover media (say 1110011) are chosen, beginning with the least significant bits. The first secret bit that is to be secluded in a pointed letter is 1. The cover text is examined heedfully from left to right due to the Czech/Latin regular direction. The first un-pointed letter in the cover-text is considered as a least significant bit, represented as "Czech" in the example. This letter '**Č**' should hold the first secret bit '1' noted by appending an extension character after the end of it. The second secret bit is '1' is appended to 'á', and the second word of the cover-text is known as 'Republic' where 'i' holds the secret bit '1' and 'a' is separated for concealing the secret bit '0'. However, this letter position cannot allow extension, forcing us to ignore it. The next possible pointed letter to be extended is 'Is a beautiful place'.

Secret Bits	**1110011**
Cover Text	**Česká republika jekrásné místo**
Steganographic Text	**Česká republik a jekrásné místo** **1 1 1 0 0 1 1**

Fig 3 Steganography example appending extensions after letters

Moreover, we can append extensions in front of and after the letters in the same document, rather than unalike paragraphs to be more particular. In this way it aggrandizes the the security and easily confounds the intruders.

VI. Conclusion

This paper presents an unaccustomed steganography approach which is applicative for the Czech language electronic superscription. This approach aids from the feature of having points more than the partial size of the text letters. The secret bit '0' is stored in the un-pointed letters and the secluded information bit '1' is stored in the pointed letters. The data needs to be in symmetry to the cover-text letters, ergo all the letters do not carry secret bits. Excess Czech extension characters are used nearby the letters for jotting the specific letters clenching the hidden secret bits. The advantage of letter extension is, there is no proclivity towards the inscription content.

During the concealed exchange of the information by substantiate stealthy communication and text documents,

ISBN: 978-988-19253-4-3
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

and the integral aspects of the steganography are essential, such as the reliability, robustness, security. This method features all these indispensable phases effectively. This method can also be applied on other languages which display coequal texts to Czech. The distinct features of this unaccustomed Czech text steganography technique exercising the letter extensions is captivating for the information security.

ACKNOWLEDGMENT

We are extremely grateful to the Head of the Department **Ms. B. Padhmavathi** for the support extended to us for this work. We are also grateful to **Mr. C. Rajivegandhi** for encouraging and helping us in this work. Without their help it would have been very hard to complete this work.

REFERENCES

[1] B. Chen and G.W. Wornell, " Quantization Index Modulation: A Class of Provably Good Methods for Digital Watermarking and Information Embedding,' IEEE Trans. Information Theory, Vol. 47, No.4, pp. 1423-1442, 2001

[2] William Stallings, "*Cryptography and Network Security Principles and Practice*", Pearson Edition India, Fifth Edition, 2012.

[3] G. Doërr and J.L. Dugelay, "A Guide Tour of Video Watermarking", Signal Processing: *Image Communication*, Vol. 18, No 4, pp. 263-282, 2003.

[4] M. Hassan Shirali-Shahreza, Mohammad Shirali-Shahreza, " A New Approach to Persian/Arabic Text Steganography," *5th IEEE/ACIS International Conference on Computer and Information Science (ICISCOMSAR 06)*, pp. 310- 315, July 2006.

[5] K. Gopalan, "Audio steganography using bit modification", Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, (ICASSP '03), Vol. 2, pp. 421-424, April 2003.

[6] J.C. Judge, "Steganography: Past, Present, Future", *SANS white paper*, http://www.sans.org/rr/papers/, November 30, 2001.

[7] R. Chandramouli, and N. Memon, "Analysis of LSB based image steganography techniques", *Proceedings of the International Conference on Image Processing*, Vol. 3, pp. 1019 – 1022, Oct. 2001.

[8] K. Bennett, "Linguistic Steganography: Survey, Analysis, and Robustness Concerns for Hiding Information in Text", *Purdue University, CERIAS Tech.* Report 2004-13, 2004.

[9] K. Rabah, "Steganography-The Art of Hiding Data", *Information Technology Journal*, vol. 3, Issue 3, pp. 245-269, 2004.

[10] W. Bender, D. Gruhl, N. Morimoto, and A. Lu, "Techniques for data hiding", *IBM Systems Journal*, Vol. 35, No 4, pp. 313-336, 1996.

[11] S.H. Low, N.F. Maxemchuk, J.T. Brassil, and L. O'Gorman, "Document marking and identification using both line and word shifting", *Proceedings of the Fourteenth Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM '95)*, April 1995.

[12] N. Provos and P. Honeyman, "Hide and Seek: An Introduction to Steganography", *IEEE Security & Privacy*, pp. 32-44, May/June 2003.

[13] M. H. Shirali-Shahreza, and S. Shirali-Shahreza, "A Robust Page Segmentation Method for Persian/Arabic Document", *WSEAS Transactions on Computers*, vol.4,Issue11,Nov.2005,pp.1692-16.