Applied Mathematics & Information Sciences
An International Journal

http://dx.doi.org/10.12785/amis/081L47

Secure Data Hiding in Binary Text Document for Authentication

Debnath Bhattacharyya[1,*], *Asmita Haveliya*[2,*] *and Tai-hoon Kim*[3*]

[1] Department of Computer Science and Engineering, FET, NSHM Knowledge Campus, Durgapur-713212, West Bengal, India
[2] Department Of Electronics, ASET, Amity University, Lucknow, India
[3] Department of Convergence Security, Sungshin Women's University, 249-1, Dongseon-dong 3-ga, Seoul, 136-742, Korea

Received: 23 May. 2013, Revised: 17 Sep. 2013, Accepted: 18 Sep. 2013
Published online: 1 Apr. 2014

Abstract: In an effort to detect a secure technique for the text document, we examined the Steganography techniques for secure data transmission. Steganography deals with the understanding of information hiding in a way that only sender and receiver can understand and realize the subsistence. It is a method of hiding the transmitted data in the form of codes that is hiding the data in image text etc, although in the proposed work we are concerned about the text data hiding in text document, and thus the author of the proposed work has used illustrations in DOC, PPT, TXT, MATLAB script file(.M) formats. There are various methods of information hiding but in text Steganography it is not easy to hide information as text data provides us less redundant space for concealing the secrets. Simulation outcomes demonstrate that data embedding based on coding procedure pursued has no traceable distortions in the host file. The results for the proposed work are obtained using the MATLAB version 7.10.0.499(R2010a).

Keywords: Steganography, hiding data, text file encryption, text file decryption, covert writing, ppt cover data, doc Steganography.TXT and MATLAB SCRIPT files cover data

1 Introduction

Before having a discussion on the subject of Steganography, let?s deal with information hiding. Information hiding consists of processes and schemes to bury or cover the information from intermediary, for instance hackers. The sender requirements is to put out of sight his data from third parties and also clearly has desires to communicate it through a technique with the intention of no more than receiver obtains it. For this purpose we require a technique and Steganography is a tool which provides the accessibility to the sender to hide from view the transmitted information. The secret message file and the cover file in which the message file is to be embedded can be text, any image, video, an audio, executable file and so forth.

Sending information clandestinely and communicating covertly encompass an immense amount of attention for epoch. These data hiding technique is exceptionally to a large extent in use from ancient times. It was as well employed by German Nazis soldiers to dispatch the messages. Invisible ink was also one method of Steganography for information hiding in ancient epoch. for the duration of World War II, undetectable invisible inks presented a common form of invisible writing. With the invisible ink, a seemingly innocent letter could contain a very different message written between the lines. Therefore, the document text can conceal a hidden meaning through the use of unfounded ciphers (unencrypted message), which absolute perfectly camouflage the genuine real message in an ordinarily appearing letter. There are quite a lot of methods for this and even more highly secured methods are under investigation and are increasing gradually. If we have a discussion on the subject of text Steganography, then we will become conscious of the fact that the chief concern in text Steganography is the redundancy of data. During recent times, in the field of text Steganography it was specifically designed and deliberated to take advantage of the specific, detailed characteristic of the target language.

Here, in the proposed work author deals not with the specific, detailed characteristic of any targeted language, but with the file types that can have text information inside like the word file with .DOC extension, PowerPoint file with .PPT extension, the notepad file with .TXT

* Corresponding author e-mail: debnathb@gmail.com, asmita.haveliya@gmail.com, taihoonn@daum.net

extension, and the MATLAB script file with .M extension. This method gives the advantage of having any language involved; it is a upgrading in the enhancement of the Steganography technique which now can be applied to a file type, which is capable to encompass whichever language. This facilitates programmer not to have knowledge of all the languages actually being worked on.

2 Literature Survey

Mohammad Shirali-Shahreza in his paper introduces the method which uses, changing the word of American and England language to hide the information. There are many words in both languages which are pronounced same but are different in spelling and this technique deals with them to give privacy to information which has to be transmitted. We can use either words or spelling for hiding, for example: “dialog” is an American word which is “dialogue” in British language and from these types of words having similar sound but different in character an efficient way of information hiding takes place which is called as text Steganography [1].

M. Hassan Shirali-Shahreza and Mohammad Shirali-Shahreza, in their paper [2] proposed to hide the information in Persian and Arabic words which have same shape but are different in meaning. Arabian and Persian are very much similar language and only have four different letters. These two languages collapse to generate the message which hides the data bits. We only have to change the character for the code which we want to hide and a stego text or coded information is generated.

Well, L. Y. POR, B. Delina in 2008 introduced approaches of text Steganography which is a part of information hiding. Text Steganography is the art of sending message hidden into texts. It is a convenient approach to hide the information that is being used in ancient times. This document deals with inter word and inter paragraph spacing to hide the data which generate a cover page dynamically [3].

Nuur Alifah Roslan, Ramlan Mahmod, Nur Izura Udzir in their documentation deals with the text Steganography in Arabic text. Arabic text Steganography uses a sharp edges method, which allows one to hide the secret bits in the sharp-edges for each character in the Arabic text document. This technology is very much dedicated to the hiding capacity of information. This technique allows hiding the information more in amount and hence very much needed that it is introduced to public also [4].

3 Proposed Work

The process of data hiding through Steganography in the proposed work pursue the following steps, first of all it reads the input cover file and the secret file that is to be embed subsequently it checks if the size of the secret file is less than the 5% of the size of the cover file, it will be encrypted. Otherwise an error message is displayed and will ask to replace the existing cover file with whichever file of large size, then user will necessitate selecting another cover file with appropriate size. As the cover file is a word file with .DOC extension, PowerPoint file with.PPT extension, the notepad file with .TXT extension, or the MATLAB script file with .M extension, therefore the necessity is to first convert the matrix into single dimension, and the secret file is also converted into single dimension. To add to the security of the covert message file, it is converted into binary format and each binary data undergoes right shift, XOR operation, then again right shift and finally XOR operation for those binary bits, Next we need to convert the above binary format secret file into single dimension. Then requirement is to convert the single dimensional cover file data into binary format, Followed by replacement of the 1st bit of the binary cover file data with the one?s complement of the binary format secret file data. Next convert the single dimensional cover file data into decimal format to obtain the stego-image. By taking an illustration this entire development can be with no trouble understood

Assume cover file =

1 6
2 7
3 8
4 9
5 10

And Secret file=10
Before Bit exchange the secret file in binary form is

1 0 1 0 0 0 0 0

Now we perform right shift, XOR operation, then again right shift and finally XOR operation on the above secret pattern of binary file and after this Bit exchange the secret file in binary form is

1 0 0 0 0 0 1 0

Take one’s complement of the secret message

1 0 0 0 0 0 1 $->$ secret message
0 1 1 1 1 1 0 1 $->$ 1’s complement

Cover file in binary format

1 0 0 0 0 0 0 0 :1
0 1 0 0 0 0 0 0 :2
1 1 0 0 0 0 0 0 :3
0 0 1 0 0 0 0 0 :4
1 0 1 0 0 0 0 0 :5
0 1 1 0 0 0 0 0 :6
1 1 1 0 0 0 0 0 :7
0 0 0 1 0 0 0 0 :8
1 0 0 1 0 0 0 0 :9
0 1 0 1 0 0 0 0 :10

Now we will embed the secret message at the 1st bit

0 0 0 0 0 0 0 0
1 1 0 0 0 0 0 0

$$\begin{matrix} \underline{1} & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ \underline{1} & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ \underline{1} & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ \underline{1} & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ \underline{0} & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ \underline{1} & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \end{matrix}$$

Thus the stego file =

$$\begin{matrix} 0 & 7 \\ 3 & 6 \\ 3 & 9 \\ 5 & 9 \\ 5 & 10 \end{matrix}$$

At the receivers end Decryption process is followed by entering the stego file. The secret file size, password and Pattern are extracted from the cover file. The user needs to enter the password. If the password mismatch, then the decryption is terminated. If the password matches, then the decryption is performed. The decryption process follows the same procedure of the encryption but in the reverse direction. The extracted secret file is stored. Then secret file is displayed.

4 Result

The results for the proposed work are investigated with samples of word file with .DOC extension, PowerPoint file with.PPT extension, the notepad file with .TXT extension, and the MATLAB script file with .M extension. In the ensuing four cases (fig 1, 2, 3, 4) of the results acquired in the projected work, the cover media is a word file with the extension .DOC while the secret messages are taken as word file with .DOC extension (fig 1), PowerPoint file with .PPT extension (fig 2), notepad file with .TXT extension (fig 3), and MATLAB script file with .M extension (fig 4) respectively.

Now in next four cases (fig 5, 6, 7, 8) of the results acquired in the projected work, the cover media is a PowerPoint file with the extension .PPT while the secret messages are taken as word file with .DOC extension (fig 5), PowerPoint file with .PPT extension (fig 6), notepad file with .TXT extension (fig 7), and MATLAB script file with .M extension (fig 8) respectively.

The next four cases (fig 9, 10, 11, 12) of the results acquired in the projected work, the cover media is a notepad file with the extension .TXT while the secret messages are taken as word file with .DOC extension (fig 9), PowerPoint file with .PPT extension (fig 10), notepad file with .TXT extension (fig 11), and MATLAB script file with .M extension (fig 12) respectively.

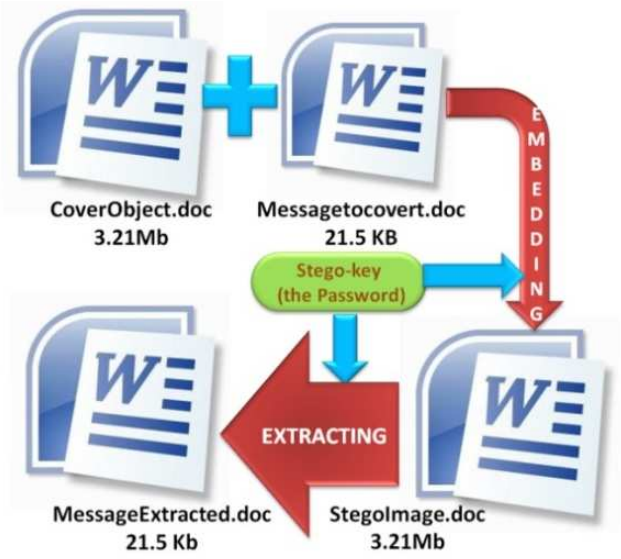

Fig. 1: Files in use for investigation : cover file of .doc extension and secret file with .doc extension

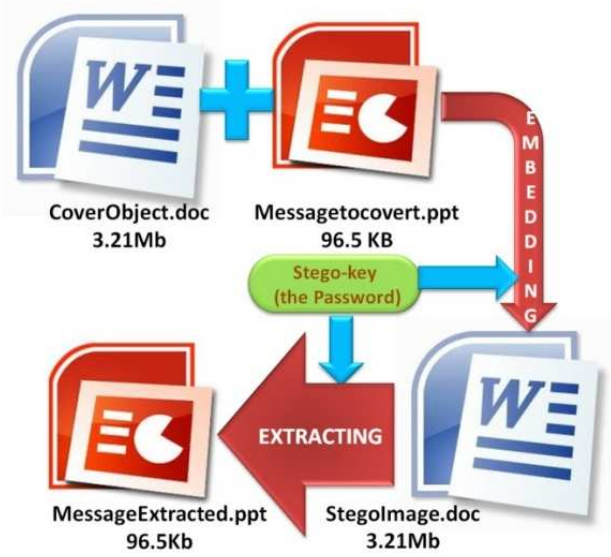

Fig. 2: Files in use for investigation : cover file of .doc extension and secret file with .ppt extension

Now in next four cases (fig 13, 14, 15, 16) of the results acquired in the projected work, the cover media is a MATLAB script file with the extension .M while the secret messages are taken as word file with .DOC

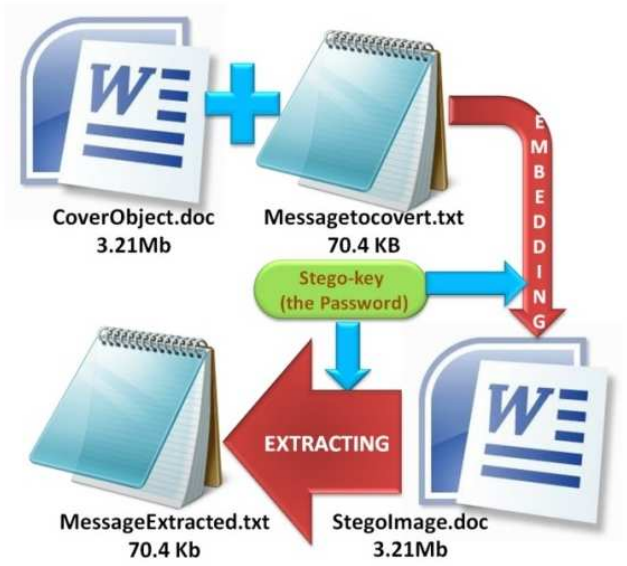

Fig. 3: Files in use for investigation : cover file of .doc extension and secret file with .txt extension

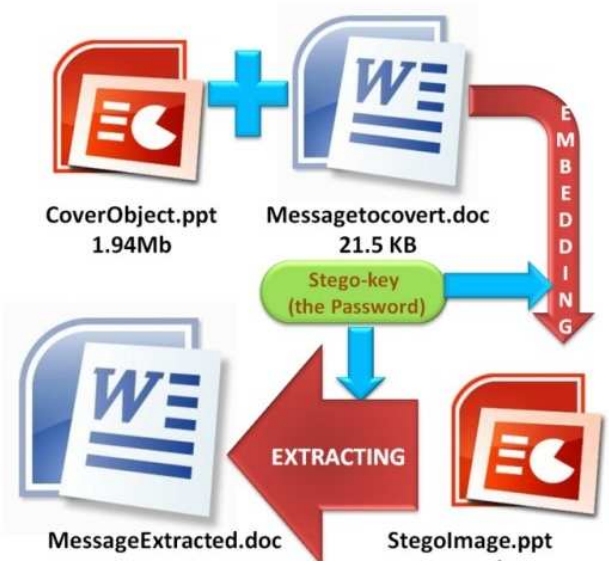

Fig. 5: Files in use for investigation : cover file of .PPT extension and secret file with .doc extension

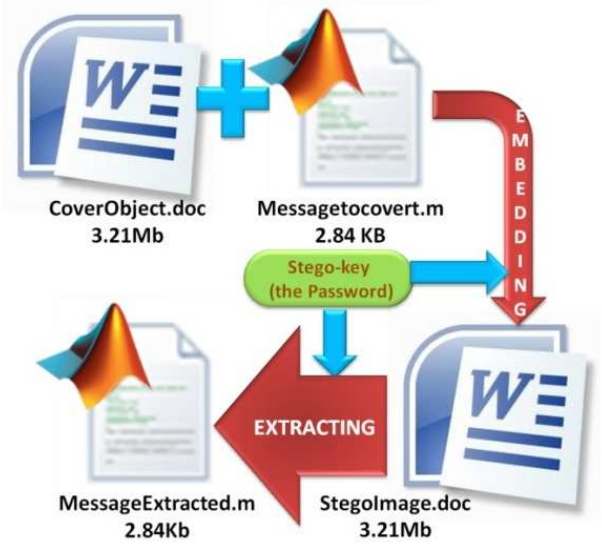

Fig. 4: Files in use for investigation : cover file of .doc extension and secret file with .m extension

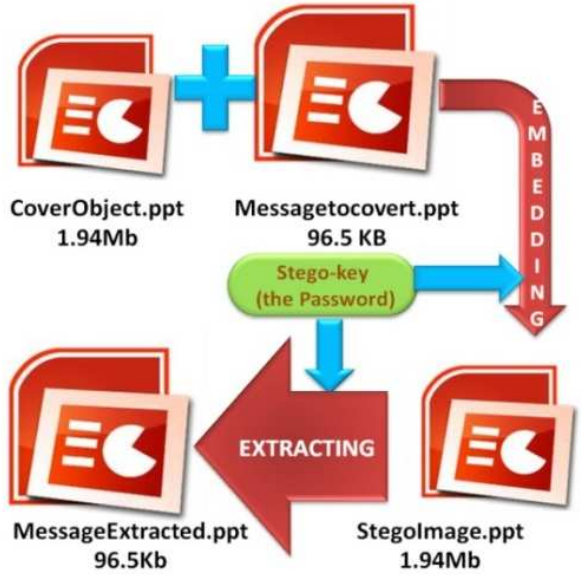

Fig. 6: in use for investigation : cover file of .PPT extension and secret file with .PPT extension

extension (fig 13), PowerPoint file with .PPT extension (fig 14), notepad file with .TXT extension (fig 15), and MATLAB script file with .M extension (fig 16) respectively.

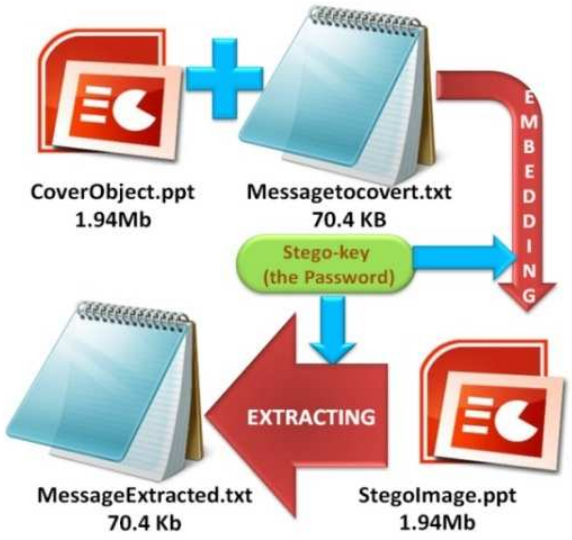

Fig. 7: Files in use for investigation : cover file of .PPT extension and secret file with .TXT extension

Fig. 9: Files in use for investigation : cover file of .TXT extension and secret file with .DOC extension

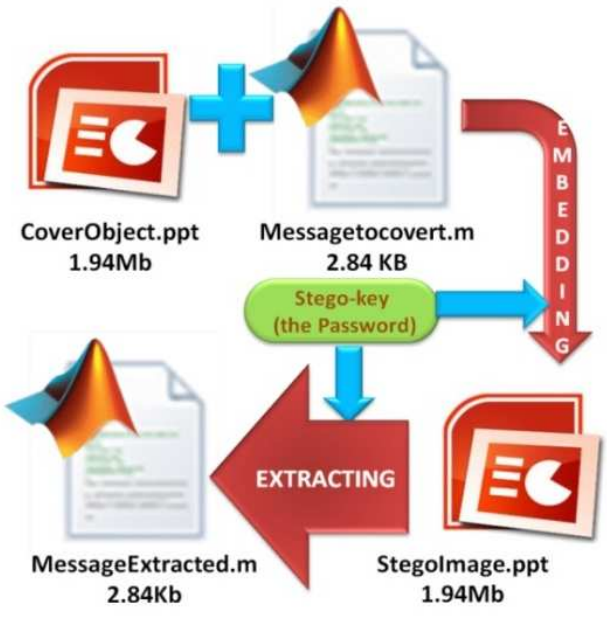

Fig. 8: Files in use for investigation : cover file of .PPT extension and secret file with .M extension

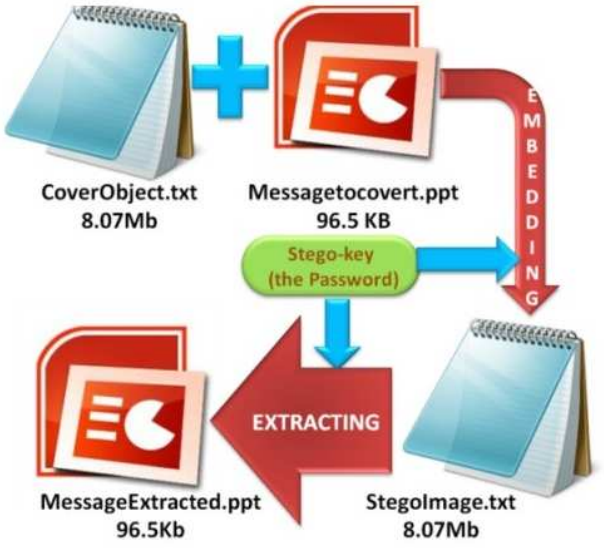

Fig. 10: Files in use for investigation : cover file of .TXT extension and secret file with .PPT extension

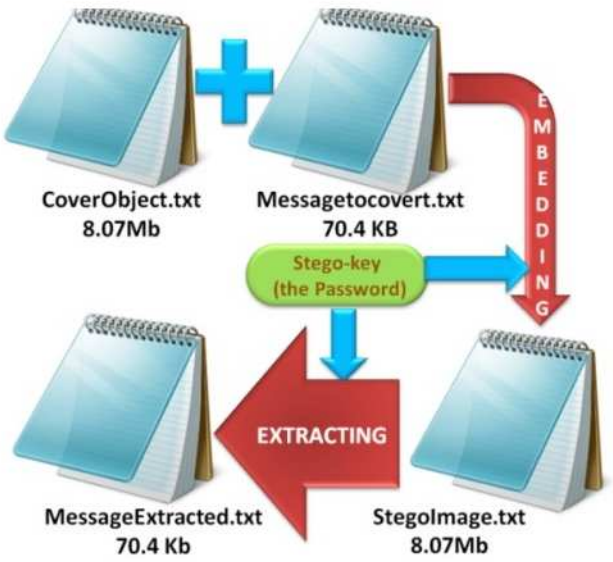

Fig. 11: Files in use for investigation : cover file of .TXT extension and secret file with .TXT extension

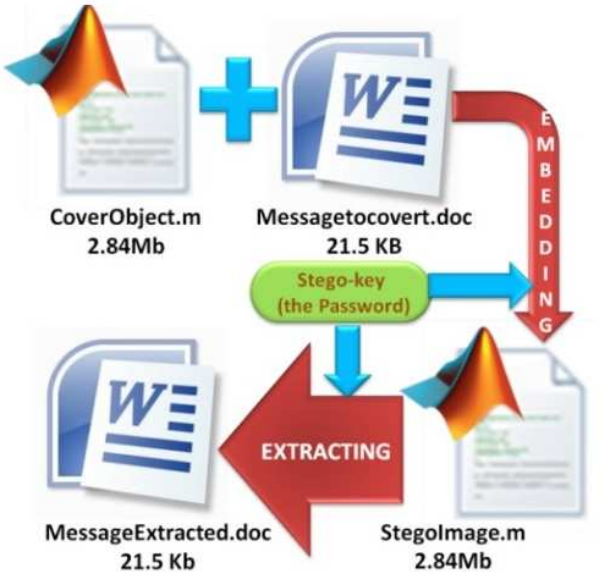

Fig. 13: Files in use for investigation : cover file of .M extension and secret file with .DOC extension

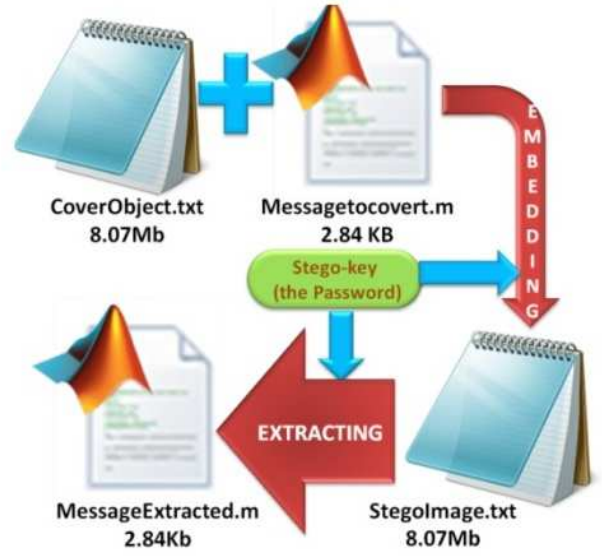

Fig. 12: Files in use for investigation : cover file of .TXT extension and secret file with .M extension

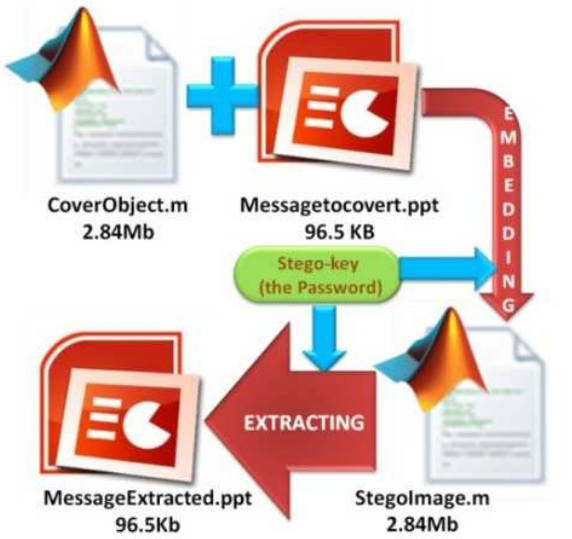

Fig. 14: Files in use for investigation : cover file of .M extension and secret file with .PPT extension

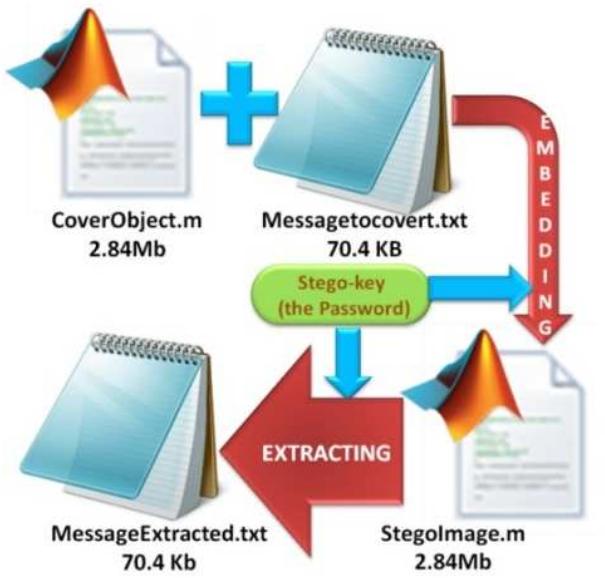

Fig. 15: Files in use for investigation : cover file of .M extension and secret file with .TXT extension

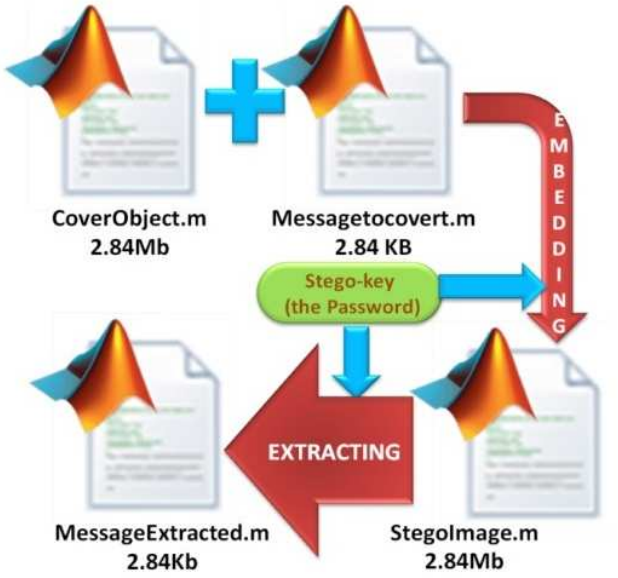

Fig. 16: Files in use for investigation : cover file of .M extension and secret file with .M extension

5 Conclusion

In the presented work, the chief essential matter is the discretion and secrecy of the buried, confidential message data being communicated. The proposed technique presents the improvement to enclose whichever language; it is a advancement in the enrichment of the Steganography practices.This algorithm harmony is with the text Steganography and is pretty tremendously dexterous as here it is tough and tricky to figure out the authenticate information that is veiled at the back. The outcomes for the intended work are attained using the MATLAB version 7.10.0.499(R2010a).

Acknowledgement

This work was supported by the Sungshin Women's University Research Grant of 2013.

References

[1] Mohammad Shirali-Shahreza, "Text Steganography By Changing Words Spelling", **2008**, 17-20 (2008).
[2] M. Hassan Shirali-Shahreza and Mohammad Shirali-Shahreza, "Arabic/Persian Text Steganography Utilizing similar Letters With Different Codes", The Arabian Journal For Science And Engineering, **35**, (2010).
[3] L. Y. Por, B. Delina, "Information Hiding: A New Approach In Text Steganography", 7th Wseas Int. Conf. On Applied Computer & Applied Computational Science (Acacos '08), Hangzhou, China, (2008).
[4] Nuur Alifah Roslan, Ramlan Mahmod, Nur Izura Udzir, "Sharp-Edges Method In Arabic Text Steganography", Journal Of Theoretical And Applied Information Technology 15th, **33**, (2011).

Debnath Bhattacharyya M.Tech (Computer Science and Engineering) from West Bengal University of Technology), Ph.D. (Tech., Computer Science and Engineering) from University of Calcutta, currently, associated with Computer Science and Engineering Department, Faculty of Engineering and Technology, NSHM Knowledge Campus Durgapur, as a Professor and Head. Dr. Bhattacharyya has 17 years of experience in Teaching and Research. He published 5 Text Books for B.Tech, and MCA, so far. He also published 135 Research Papers in International Journals and Conferences. His Research Interests include Biometric Recognition, Pattern Recognition and Image Processing. He is also associated with West Bengal University of Technology, University of Calcutta and many leading National and International Universities as the Ph.D. Supervisor. He is a Member of IEEE, IACSIT and CSI. He conducted many International Conferences as a General Chairs and delivered Keynote Speeches in many International Conferences.

Asmita Haveliya M.Tech (Electronics and Communication Engineering) from Amity University, Lucknow. She is currently associated with ALH Academy, Lucknow. Her research interest includes Data Hiding and Authentication. She published 10 Research Papers in International Journals and Conferences.

Tai-hoon Kim received his M.S. degrees and Ph.D. in Electrics, Electronics & Computer Engineering from the Sungkyunkwan University, Korea. And he got his 2nd Ph.D. in School of Information and Computing from University of Tasmania, Australia. After working with Technical Institute of Shindoricoh 2 years as a researcher and working at the Korea Information Security Agency as a senior researcher 2 years and 6 months, he worked at the DSC (Defense Security Command) about 2 years. After working with Hannam University four and a half year as an associate professor, now he is currently working at Sungshin W. University. He wrote 17 books about the software development, OS, and computer hacking & security. And he published about 200 papers by 2012.He is a member of IEEE, ACM, KIIT and SERSC. He was a General Chair or Program Committee chair from many international conferences.