Secure PDF Text Steganography by Transforming Secret Into Imperceptible Coding

Sanjive Tyagi, Rakesh Kumar Dwivedi, Ashendra Kumar Saxena

Abstract*: In this paper, we have proposed an approach of PDF based text Steganography by considering the hiding capacity and security of secret information, and improved imperceptibility of stego-cover file. In proposed approach secret information are transformed into compact and encrypted form of imperceptible coding, and then translated into bits form, thereafter embedded into targeted locations of PDF file by applying new cross-reference coding technique in incremental updates of PDF file with less computation complexity. The proposed extraction process provides authentication of received stego-cover file such that only desired file is accepted for the extraction process otherwise fake file is discarded by recipient. Time complexity has been improved significantly by implementing a novel method of PDF steganography for embedding the secret data. Experimental result demonstrates that proposed method provides efficient algorithms in terms of improved security of hidden information.

***Keywords*: Compatibility, Cross-reference, Distortion, e-document, Incremental, Steganography.**

I. INTRODUCTION

Since the World Wide Web has been introduced applicability of internet, personal computers, android mobile phones, transmission of digital documents and various digital format files growing exponentially. This explosive growth of digital documents generates the need on compatibility of digital documents within complex file formats of e-text processor, e-power point presentations, e-spreadsheets, e-graphics designers so on. Document processing software are used to manipulate, compose, modify, formatting, styling of document to make it attractive and informative for the purpose of reading, communicating, documenting the textual records. There are various document processing software available across the world on different platform which are not compatible on cross-platform environment so need of platform-independent textual record format was there, which can be used for the e-exchange of documents over cross-platform environment. A PDF (Portable Document Format) was developed as Acrobat products in 1993 for creating the platform-independent e-document by Adobe [1]. Nowadays PDF document is most popular e-document format for cross-platform documents exchange through the internet because of its qualitative advantageous features, better printing quality, and fast communication over internet

Manuscript received on January 02, 2020.
Revised Manuscript received on January 15, 2020.
Manuscript published on January 30, 2020.
* Correspondence Author

Sanjive Tyagi*, College of Computing Sciences and Information Technology, Teerthanker Mahaveer University, India, Mob-91-9219869447, E-mail: tosanjive@gmail.com

Prof. (Dr). Rakesh Kumar Dwivedi, Principal-College of Computing Sciences and Information Technology, Teerthanker Mahaveer University, India, E-mail: principal.computers@tmu.ac.in

Dr. Ashendra Kumar Saxena, Associate Professor- College of Computing Sciences and Information Technology, Teerthanker Mahaveer University, India, E-mail: drashendra.computers@tmu.ac.in

A. Introduction of Steganography

Steganography is technology of concealing secret digital information by inserting it into another digital file known as cover file with stego-key and make secret information imperceptible and obtained digitized stego-cover file can be transmitted or stored. Intended receiver can extract secret digital information from stego-cover file with stego-key by applying reverse of embedding procedure “submitted for publication” [2]. Steganography mainly required two types of digital files i.e. cover file and secret file which may be image, video, audio, text, pdf etc. whereas type of cover file depends upon the type of steganography that may be image steganography or audio steganography or text steganography or other type of steganography. The three basic ideas behind robust steganography are embedding capacity, invisibility and imperceptibility. There is a tradeoff among these features to obtain high-quality steganography [3-6].

B. Basic View of PDF Layout

A PDF file is the combination of text, graphics, and binary information. It is a set of several indirect objects broadly classified in three sections as header, body, cross-reference section and trailer, which consists of placeholders to embed secret information without affecting the appearance of PDF file [1]. Each section of PDF document is managed by page formatting description script known as PostScript [7]. First section is header consists of comments, second section consists of text objects like Tm, Td, TD operators manages the text position, operator Tj, operator Tc for character positioning, operator, Tw for managing words, operator TJ specify the layout of PDF file which independent of any platform, so on. This section is responsible for the appearance of textual records in PDF files. The third section consists of cross-reference section and trailer. The cross-reference information is arranged at end of PDF file used to locate any page quickly not dependent on the length of document consists of any number of pages. It is noticed in study that there are large number of place holders to conceal the secret information in second and third section specified in cross-reference section.

C. PDF Text Steganography

Text steganography is useful to hide the small size of secret information. In [7] author suggested hiding capacity of secret information inside cover text file is less because text has few noises in comparison to image. There are various methods of text steganography by managing the between-word space as one between-word space may denote bit 0 and two between-word space denote bit 1 etc.,

by controlling syntactic rules of language, by controlling semantic rules agreed between sender and receiver, by creating simulating text containing secret words placed under specific rules so on.

II. LITERATURE REVIEW

In [8] author proposed (t, n) threshold secret sharing scheme using cryptographic hash function to verify the shared data by authorized shareholder using Shamir's secret sharing scheme [9]. The t number of participants out of n verifies their received share using hashing and master key, which is generated at recipient side to compute shareholder key which helps in constructing the shares.

In 10] author proposed an approach to split secret information on the basis of secret key. It is noticed in study splitting should be dependent on number of shareholder of secret information.

In [11] proposed zero distortion technique in which locations of bit value of secret information are searched in cover file and their locations are stored. Creating locations of each bit of secret bit value is very much time consuming, it is fine for small size secret information but not suitable for larger secret data because time complexity is high.

In [12] author introduced various short message techniques to improve the amount of secret text for the purpose of steganography. They have proved that short message service provides a secured hiding scheme but in this scheme both sender and receiver should have knowledge of proposed online short text technique.

In [13] author proposed a PDF steganography by embedding the secret information into white space symbols such as spaces, carriage return, form feed, and null. In proposed approach hiding capacity of secret information is dependent on mentioned white spaces available in PDF file, however, few symbols are not available generally, and this is the limitation of this approach.

In [14], author proposed three methods to conceal the secret information into PDF file. They suggested incremental updates used to manage the modification of PDF file may be utilized to implant the secret information for hidden transmission over large network. We proposed a new approach of steganography and make use of a new cross-reference section and trailer based on Hon [14]. Our proposed technique provides high embedding capacity y with more security.

In [15], the author recommended PDF text steganography by encrypting every character of confidential message by Huffman coding calculation subsequently empty A0s are embedded to disguise secret characters at between-characters places of words inside the carrier PDF document.

A semantic text message steganography recommended by [16] that utilizing identical word substitution technique to conceal bit casing of secret content. Here substitution strategy is changing the original content which may conspicuous as unnatural by the human onlooker. In [17], the creator displayed text-based steganography based on LZW compression technique with email-ids and email messages as targeted carrier positions of confidential information. In [18], message steganography displayed by joining the transition to-front, tunnels wheeler-change and LZW compression techniques to acquire improved compression capacity of covert content while the carrier medium is alike [17]. In [19], an augmentation of message steganography [8] is exhibited so as to improve implanting limit by the utilization of the Huffman calculation. In [17-19] plans spread content picked is not common content that using the arrangement of email-ids as a carrier medium may make defenselessness which might be perceptible as unnatural by the human spectator and can be tempered.

A methodology of text base steganography recommended by [20] using a collection of email-ids as targeted carrier positions for hiding secret message. Moreover, in the second strategy for [21] classified bits are embedded into email body by using the shading concealing plan. These strategies utilize a lot of email-id as carrier media additionally changing the presence of original cover substance by color shading steganography scheme which prompts defenselessness for steganalysis assault. Although, our proposed strategy admirably against such errors.

The strategy in [21] presented a procedure by applying a homomorphic encryption technique with secret content. This methodology has confinement that carrier medium should comprise of encoded characters of secret content which must be implanted through our proposed work does not have such limitations. In [22], the author proposed a steganography plot in which content of host carrier document and content of secret document is produced by using the procedure of Markov chain. This reliance of carrier substance with privileged insights content leads defenselessness to steganalysis assault.

A text base steganography strategy was presented in [23] that utilizing alteration of Bengali language message for disguising private contents. They guarantee this strategy is secured against steganalysis ambushes; although, the change of cover medium substance is consistently against the impalpability.

The PDF steganography is being used to hide the confidential information within the text of PDF file. PDF document is most widely used e-document in which postscript descriptive language is used to manage the texture and graphical matters of PDF file [1]. Due to the popularity of PDF file over World Wide Web, it is suitable media to be used as cover file in text steganography.

In traditional text, steganographic techniques, some sort of adjustment happens, for example, substitution by comparable beginning, wordlist, elective word, extraordinary character, shading concealing schemes and so on, to the substance of the carrier medium which may distinguish as unnatural by a human onlooker, which prompts defenselessness for steganalysis assault. To beat this issue undetectable control ASCII code A0s are recommended to encrypt by transforming secret into imperceptible encoding and insert the secret information secretly inside PDF carrier document. Moreover, incremental updates feature of PDF file within cross-reference section are being utilized as placeholders for embedding secret information which improves the installing limit and intangibility.

The proposed cross-reference coding technique likewise reduces the overheads of stego-cover content with an adequately high payload. The PDF content steganography is being created as proposed paper to cover the secret information inside PDF content medium without influencing the visual nature of PDF document..

III. PROPOSED SCHEME

A. Authentication Process of StegoPDF File

Before extraction of secret data, authentication takes place for StegoPDF file by intended recipient, if verification is positive then only extraction process sustains otherwise StegoPDF file is discarded and request is done for another StegoPDF file.

Let us say dealer generate StegoPDF file sc_i when recipient r_i received his share, authenticate by computing hash function value $H(sc_i)$ as follow

Algorithm to Authenticate StegoPDF file sc_i by recipient r_i

```
1. Recipient get the public key
2. Compute $v_i = H(sc_i)$ using public key
3. Compute $e_i$= extracted value from ith block of
   received StegoPDF file $sc_i$
4. If $v_i$=$e_i$ then accept the StegoPDF file's share $sc_i$ by
   recipient $r_i$
5. else if $v_i$ !=$e_i$ then discard StegoPDF file's share $sc_i$
   and request to resend the valid StegoPDF file's share
   $sc_i$
6. End
```

B. New Cross-Reference Coding Technique

During conversion of number system to another form of number system, there was an application of Binary convertor. [24].

Let us demonstrate New Cross-Reference Coding Technique.

New Cross-Reference coding technique	
Steps	**Experiments Results**
Secret Data	secret information invisible
Convert into Binary Form	01101001 01101110 01100110 01101111 01110010 01101101 01100001 01110100 01101001 01101111 01101110 00100000 01101001 01101110 01110110 01101001 01110011 01101001 01100010 01101100 01100101

Segment of 16 bits, if less than 16 bits padded 0s at left side	0110100101101110 0110011001101111
	0111001001101101
	0110000101110100 0110100101101111
	0110111000100000
	0110100101101110 0111011001101001
	0111001101101001
	0110001001101100 0000000001100101

Convert 16 bit segment into decimal into number	26990 26223 29293
	24948 26991 28192
	26990 30313 29545
	25196 101

Add 10000 to get similar number digits decimal number	36990 36223 39293
	34948 36991 38192
	36990 40313 39545
	35196 10101

New cross-reference section's locations of CoverPdf file to embed decimal numbers	xref 0 14 0000000010 65535 f 0000000000 00000 n 0000000000 00000 n 0000000000 00000 n 0000000000 00000 n 0000000000 00000 n 0000000000 00000 n 0000000000 00000 n 0000000000 00000 n 0000000000 00000 n 0000000000 00000 n 0000000000 00000 n 0000000000 00000 n 0000000000 00000 n trailer
After embedding the decimal values of secret data into new cross-reference section and obtain Stego-cross-reference section.	xref 0 14 0000000010 65535 f 0000000000 00000 n 0000036990 00000 n 0000036223 00000 n 0000039293 00000 n 0000034948 00000 n 0000036991 00000 n 0000038192 00000 n 0000036990 00000 n 0000040313 00000 n 0000039545 00000 n 0000035196 00000 n 0000010101 00000 n 0000000000 00000 n trailer
Extraction Process	

Extracted secret decimal from Stego-cross-reference	36990 36223 39293
	34948 36991 38192
	36990 40313 39545
	35196 10101

Subtract 10000 from each extracted decimal number	26990 26223 29293
	24948 26991 28192
	26990 30313 29545
	25196 101

Convert each into 16 bit binary number segments	0110100101101110 0110011001101111
	0111001001101101
	0110000101110100 0110100101101111
	0110111000100000
	0110100101101110 0111011001101001
	0111001101101001
	0110001001101100 0000000001100101

Grouping each into 8 bit segment	01101001 01101110 01100110 01101111
	01110010 01101101
	01100001 01110100 01101001 01101111
	01101110 00100000
	01101001 01101110 01110110 01101001
	01110011 01101001
	01100010 01101100 00000000 01100101

Extracted secret Data	secret information invisible

Fig. 1. Demonstrating New Cross-Reference Coding Technique

C. Embedding Procedure using New Cross-Reference Coding Technique

Incremental updates of PDF file is one of the additional features of PDF file available after version 1.4. In this approach, if fresh modification in PDF's object takes place then that is managed without altering the existing objects. It is found in study that latest updates in PDF objects are maintained by appending new cross-reference with new trailer is added hence by appending a new cross-reference section, that can be used as a carrier of secret information without impact on visibility of text of PDF file.

In our experiment new cross-reference section in Fig. 2 as destination location of cover PDF file C_i and secret data S_i as follows

```
xref
0 14
0000000010 65535 f
0000000000 00000 n
0000000000 00000 n
0000000000 00000 n
0000000000 00000 n
0000000000 00000 n
0000000000 00000 n
0000000000 00000 n
0000000000 00000 n
0000000000 00000 n
0000000000 00000 n
0000000000 00000 n
0000000000 00000 n
0000000000 00000 n
trailer
<</Size 10/Root 1 0 R/Info 9 0
R/ID[<654073D4C03C3F45BD63BCC1A23FE973><6540
73D4C03C3F45BD63BCC1A23FE973>] >>
xref
0
%%EOF
```

Fig. 2. Destination location of new cross-reference section

D. Steps for Embedding Secret Data S into New Cross-Reference Section of cover PDF file C

Step 1:

In order to conceal Secret Data $S= \{t_1, t_2, t_3, t_{4\ldots\ldots}t_n\}$ in new cross-reference section of cover PDF file C named CoverPdf , Secret data consists of n number characters as decided by sender. Insert m number of stream objects and append new cross-reference section and trailer, where m is number of transformed decimal values to be hidden.

Step 2:

Transform secret data $S= \{t_1, t_2, t_3, t_{4\ldots\ldots}t_n\}$ into set of decimal number where each character is considered to be represented by eight bit ASCII code, accordingly transform each character of secret data into equivalent decimals.

$\text{Decimal}(S) = \text{Decimal}\ (t_1, t_2, t_3, t_{4\ldots\ldots}t_n)$

$\text{Decimal}(S) = \{\text{Decimal}(t_1), \text{Decimal}(t_2), \text{Decimal}(t_3), \text{Decimal}(t_4),\ldots\ \text{Decimal}(t_n)\}$

For each character of $S= (t_1, t_2, t_3, t_{4\ldots\ldots}t_n)$

$\text{Decimal}(S) = \{d_1, d_2, d_3,\ldots\ldots, d_n\}$

Denoting decimal values for data S as given

$Dt_i = \text{Decimal}\ (t_i)$ where $1<= i <= n$

Step 3:

Transform each decimal number of Dt_i obtained from step-2 into 8 bits segment of binary representation using ASCII code.

$\text{binform}(Dt_i) =\{ \text{binform}\ (b_q, Dt_i) \}$

where b_q is corresponding binary value of each decimal number of secret data Dt_i, b_q represent 8 bits segment of binary number i.e. 2^8 and $1 <= i <=n$.

$\text{binform}(Dt_i)= \{b_{q1}, b_{q2},\ldots\ldots,b_{qn}\}$

these are n blocks of 8 bits based on number of characters in secret data S.

Denoting binary form for data S as given

$Bt_i= \text{binform}(Dt_i)$ where $1<= i <= n$

Step 4:

Concatenating the adjacent pair of b_{q1} & b_{q2}, b_{q3} & b_{q4}, …. so on of Bt_i obtained from step-3 in order to make segment of 16 bits, if pair is not available then padded 8 zero bits at left side of unpaired 8 bits and then transform each 16 bits segment into an equivalent decimal number

$\text{Decimal}\ (Bt_i) =\{\text{desiform}(b_q, Bt_i)\}$

where b_q is 16 bits segments of Bt_i, b_q represent 16 bits segment of binary number i.e. 2^{16} and $1 <= i <=n/2$.

$\text{Decimal}\ (Bt_i) = \{D_1, D_2, D_3,\ldots\ldots, D_{n/2}\}$

Denoting obtained decimal form for data S as given

$DBt_i = \{\text{decimal}(16\ \text{bits}\ (Bt_i)) \}$

Step 5:

Transform DBt_i into matrix HideDBt[nRow][nCol] where each element of matrix is decimal value obtained from step-4.

HideDBt[nRow][nCol] = $\{DBt_i\}$ where $1<= i <=n/2$, nCol=16, nRow=n/32.

Step 6:

Convert each cell of HideDBt[nRow][nCol] into equal number digits of decimal numbers by adding number max, where max is determine as given

if maximum number digits in decimal numbers of $\{ DBt_i\}$ is 2 then max = 10

else if maximum number digits in decimal numbers of $\{DBt_i\}$ is 3 then max = 100

else if maximum number digits in decimal numbers of $\{DBt_i\}$ is 4 then max = 1000

else if maximum number digits in decimal numbers of $\{DBt_i\}$ is 5 then max = 10000

so on

Step 7:

Each cell of HideDBt[nRow][nCol] obtained from step-4 is embedded into trailer of new cross-reference section of cover PDF file named CoverPdf using Incremental Updates approach of PDF file.

Input: Cover PDF file C named CoverPdf, Secret data HideDBt[nRow][nCol], nCol=16, nRow=n/32.

Output: CoverPdfStego

Algorithm for embedding

1. Open CoverPdf PDF file in binary stream form on read and write mode
2. Locate trailer of cross-reference section in CoverPdf file
3. Insert new trailer and new cross-reference section
4. Insert stream object into trailer of cross-reference section

```
5. BOF ← 01 // represents ASCII code for beginning of
     secret data
6. Embed BOF into 10 digit number of stream of object //to
     mark beginning of secret data
7. Update 10 digit number of stream object in trailer of
     cross-reference section of coverPdf file
8. nRow ← n/32   // obtained from step-5 above this
     algorithm
9. nCol ← 16 // obtained from step-5 above this algorithm
10. Transform secret data obtained from step-6 above of
     the algorithm into matrix HideDBt[nRow][nCol]
11. CtrRow ← 1
12. while (CtrRow < = nRow)
13.         CtrCol ← 1
14.         while (CtrCol < = nCol)
15.             Insert new stream object in trailer of cross-
                  reference section
16.             Embed HideDBt[CtrRow][CtrCol] into 10
                  digit of number of stream object
17.             Update 10 digit number of stream object in
                  trailer of cross-reference section of
                  coverPdf file
18.             CtrCol ← CtrCol +1
            endwhile
19. CtrRow ← CtrRow + 1
20. endwhile
21. Insert stream object into trailer of cross-reference //to
     mark end of secret data
22. EOF ← 03 // represents ASCII code for End of secret
     data
23. Embed EOF into 10 digit number of stream of object
24. Update 10 digit number of stream object in trailer of
     cross-reference section of coverPdf file
25. Close CoverPdf file
26. Rename CoverPdf to CoverPdfStego
27. End
```

E. Steps for Extracting of Secret Data from New Cross-Reference Section

Step 1:

Locate the trailer section where secret data is hidden in CoverPdfStego PDF file by searching BOF which represented by ASCII code 01 denoting beginning of secret data marked at the time of embedding, if it is found then proceed the extraction and store them into new file SecretData.

Input: Stego-Cover PDF file named CoverPdfStego.

Output: Secret Data file named SecretData in the form of decimal value.

Algorithm

```
1. Open CoverPdfStego PDF file in binary stream form on
     read mode
2. Open SecretData file on write mode
3. Locate trailer of cross-reference section in CoverPdfStego
     file
4. Locate BOF in stream object in trailer of cross-reference
     section if found then locate next 10 digit number of
     stream object and store it into variable VarSd
5. else exit
   endif
6. while not eof ( CoverPdfStego )
7.     Extract secret data from VarSd and store it into
       EVarSd
8.     If EVarSd=EOF then break while loop
9.     else
10.    write EVarSd into SecretData file
11.    write " " into SecretData file // create a space
       between two decimal numbers
       endif
12.    locate next 10 digit number of stream object and
       store it into variable VarSd
13. endwhile
14. Close SecretData file
15. Close CoverPdfStego file
16. End
```

Step 2:

A secret data file named SecretData is obtained from previous step consists of decimal number have to be converted into readable form secret information. Transform SecretData file into ASCII characters of readable form as follows

Algorithm

Input: Secret Data file SecretData in the form of decimal value, max.

Output: ExtractSecretData file consists of ASCII characters corresponding to decimal numbers.

```
1. Open SecretData file on read mode
2. Open ExtractSecretData file on write mode
3. Locate first decimal number in SecretData file
4. Read located first decimal number and store it into
   VarSecretInfo
5. while not eof(SecretData)
6.    VarSecretInfo ← Subtract the max from
      VarSecretInfo //describe in step 4 of embedding
      process
7.    Compute VarSecretInfo ←16 bits binary number
      of VarSecretInfo
8.    VarSecretInfo ←Arrange in two segment of 8 bits
9.    VarSecretInfo ←Get two decimal number from
      two segments of VarSecretInfo
10.   VarSecretInfo ← Get two ASCII character from
      two decimal number of VarSecretInfo
11.   write VarSecretInfo into ExtractSecretData file
12.   Locate next decimal number in SecretData file
13.   Read located decimal number and store it into
      VarSecretInfo
14.       if eof (SecretData)
15.           break
16.       endif
   endwhile
17. Close SecretData
18. Close ExtractSecretData
19. End
```

F. Information Hiding

Objective of this approach is to show that new cross-reference section contains suitable locations to hide secret information without distortion in quality of text of Stego-cover PDF file.

A secret data file named “Secret Message.txt” containing 28 characters and size 28 bytes and cover PDF file named “Sample Text Cover File.pdf” containing 333 characters as a text and size is 81,183 bytes. The text of “Secret Message.txt” have to be concealed into “Sample Text Cover File.pdf” and obtained “StegoCover PDF File.pdf” containing 333 characters as a text and size 81,463 bytes by applying proposed algorithm “Embedding Secret Data into New Cross-Reference Section of cover PDF” describe above.

An extracted secret data file named “Extracted Message.txt” containing 28 characters, size 28 bytes and Stego PDF file named “StegoCover PDF File.pdf” containing 333 characters as a text and size 81,463 bytes.

The text of “Extracted Message.txt” containing 28 characters is extracted from “StegoCover PDF File.pdf” by applying proposed algorithm “Extracting of Secret Data from New Cross-Reference Section” describes above.

G. Experimental results and discussion

The proposed PDF Steganography scheme evaluated by embedding secret data into "Sample Text Cover File.pdf" file by implementing our proposed embedding algorithm and embedded data in cover PDF file is may be shown and verified in UltraEdit editor. Proposed PDF based text steganography compared with image based steganography in term of hiding capacity by measuring number of bits required of cover file to embed one secret data bit presented in Table 1.

S N	Size of PDF cover File(bytes)	Size of secret data (bytes)	Size of PDF Stego-cover file (bytes)	Increase in size of Cover-Stego PDF File (bytes)	Increase in size (byte) of Stego-cover PDF file to hide to one byte	Increase in size (byte) of Stego-cover PDF file to hide to one bit
1	81,183	28	81,463	280	10	1.25
2	81,183	56	81,743	560	10	1.25
3	81,183	84	82,023	840	10	1.25
4	81,183	112	82,303	1,120	10	1.25
5	81,183	140	82,583	1,400	10	1.25

Table 1 Performance analysis in term embedding capacity of text PDF steganography

S N	Image Steganography technique	Hidden capacity (bits/pixels)	Bits hidden per 32 pixels	Pixels needed per bit	Hidden capacity per 4 byte
1	Su Q et al.'s scheme	0.03125	1	1 bit is hidden per 32 pixels	1 bit per 4 byte
2	Mansi S. Subhedar et al.’s scheme	0.25	8	8 bit is hidden per 32 pixels	8 bits per 4 byte

Table 2 Embedding capacity of image based steganography

Embedding capacity of proposed PDF based text steganography ensures better performance because 1.25 bytes needed to hide one bit of secret data that exhibits improved capacity with zero distortion in quality of PDF text. If we compare with image steganography it is better than Su et al.'s scheme [25] and may be improved to Mansi S. Subhedar et al.’s technique [26] by compressing the secret text before implanting it into cover PDF file. The difference in size of cover PDF file and Stego-cover PDF file indicates better performance in terms of distortion, no suspicious so that less vulnerable to attack.

IV. CONCLUSION AND FUTURE WORK

It is found from presented study that proposed PDF based text steganography works against statistical attack and visual attack due to better visual quality with appropriate payload. Most of existing text steganography changes the statistical properties of carrier PDF file during the implanting of covert information, whereas our approach provides no changes in visual quality of text of Stego-PDF file. Proposed technique ensures high security by embedding the secret information in encrypted form into new cross-reference section where no one can doubt about embedded data because there is no deviation in quality of text of PDF file; therefore detection of secret data is not possible. Authentication of received Stego-cover files is achieved towards recipient side that makes it more reliable in order to obtain error-free extraction of secret information.
Future research may apply this text steganography technique to conceal an image by converting it into segments of bits and then encode them into encrypted form and, then implement the proposed approach to embed confidential image into cover PDF file.

REFERENCES

1. Adobe Systems Incorporated. PDF Reference Sixth edition, Version 1.7, November 2006, http://www.adobe.com/content/dam/Adobe/en/devnet/acrobat/pdfs/pdf_reference_1-7.pdf.
2. S. Tyagi, R. K. Dwivedi, A. K. Saxena, 2019, “A Novel Data Hiding Tool Based on Pixel-Value Differencing: SteganoPixTrans”, *International Journal of Scientific & Technology Research*, (IJSTR-SCOPUS Indexed).
3. S. Tyagi, R. K. Dwivedi, A. K. Saxena, 2019, “High Capacity Steganography Protected using Shamir’s threshold scheme and Permutation Framework”, *International Journal of Innovative Technology and Exploring Engineering* (IJITEE- SCOPUS Indexed), Vol.8, No. 9S, July 2019, DOI: 10.35940/ijitee.I1127.0789S19
4. S. Tyagi, R. Kumar Dwivedi, A. K. Saxena, (2019), "A High Capacity PDF Text Steganography Technique Based on Hashing Using Quadratic Probing", *International Journal of Intelligent Engineering and Systems* (IJIES-SCOPUS Indexed), Vol.12, No.3, 2019, DOI: 10.22266/ijies2019.0630.20.
5. S. Tyagi, R. K. Dwivedi, A. K. Saxena, (2019), "A Novel PDF Steganography Optimized Using Segmentation Technique", *International Journal of Information Technology*, Springer, DOI 10.1007/s41870-019-00309-7.
6. S. Tyagi, A. K. Saxena, S. Garg, “Secured High Capacity Steganography using Distribution Technique with Validity and Reliability”, *in Proceedings of International Conference on System Modeling & Advancement in Research Trends*, Moradabad, India, pp.109 –114, 2016.
7. J.-T. Wang and W.-H. Tsai, “Data hiding in PDF files and applications by imperceivable modifications of PDF object parameters”, *Information Networks (II) – Sub-project 5: Network Security with Project No. NSC-96-2752-E- 009-006-PAE.*
8. K. M. Faraoun, “A Novel Fast and Provably Secure (T, N)-Threshold Secret Sharing Construction for Digital Images”, *Journal of Information Security and Applications*”, (2014), Vol. 19, No. 6, pp. 331-340.
9. A. Shamir, “How to Share a Secret”, *Communication of the ACM*, Massachusetts Institute of Technology, Vol. 22, pp. 612-613.

10. S. Gael, R. Ekodeck, R. Ndoundam, "PDF steganography based on Chinese Remainder Theorem", *Journal of information security and applications* (2016), Elsevier, Vol. 29, 2016, pp. 1-15.
11. Shivania,Virendra Kumar Yadava, Saumya Bathamb, "A Novel Approach of Bulk Data Hiding using Text Steganography", *Procedia Computer Science*, 57 (2015) 1401 – 1410,Published by Elsevier.
12. W. Ren, Y. Liu, J. Zhao, "Provably Secure Information Hiding via Short Text in Social Networking Tools", *Tsinghua Science And Technology*, ISSN 1007-0214 1/18 pp225-231, Volume 17, Number 3, June 2012.
13. Y. C. Lai and W. H. Tsai, "Covert Communication via PDF Files by New Data Hiding Techniques", In: *Proc. of Conf. on Computer Vision, Graphics and Image Processing*, Taiwan, China, 2009.
14. H. Liu, L. Li, J. Li, and J. Huang, "Three Novel Algorithms for Hiding Data in PDF Files Based on Incremental Updates", *eBook* of *Digital Forensics and Watermarking,* Vol. 7128, pp.167–180, 2012.
15. I. S. Lee and W. H. Tsai, "A New Approach to Covert Communication Via PDF Files", *Signal Processing*, Vol. 90, No. 2, pp. 557-565, 2010.
16. S. Mahato, D. A. Khan, and D. K. Yadav, "A Modified Approach to Data Hiding in Microsoft Word Documents by Change-Tracking Technique", *Journal of King Saud University – Computer and Information Sciences*, In Press, 2017.
17. E. Satir and H. Isik, "A Compression-Based Text Steganography Method", *Journal of Systems and Software*, Vol. 85, No. 10, pp. 2385-2394, 2012.
18. R. Kumar, S. Chand, and S. Singh, "An Email Based High Capacity Text Steganography Scheme using Combinatorial Compression", In: *Proc. of the International Conf. on Confluence- The Next Generation Information Technology Summit*, Noida, India, pp. 336–339, 2014.
19. R. Kumar, S. Chand, and S. Singh, "A High Capacity Email Based Text Steganography Scheme using Huffman Compression", In: *Proc. of the International Conf. on Signal Processing and Integrated Networks*, Noida, India, pp. 53-56, 2016.
20. A. Malik, G. Sikka, and H. K. Verma, "A High Capacity Text Steganography Scheme Based on LZW Compression and Color Coding", *Engineering Science and Technology, an International Journal*, Vol. 20, No. 1, pp. 72-79, 2017.
21. N. Naqvi, A. T. Abbasi, R. Hussain, M. A. Khan, and B. Ahmad, "Multilayer Partially Homomorphic Encryption Text Steganography (MLPHE-TS): A Zero Steganography Approach", *Wireless Personal Communications*, pp. 1-23, 2018.
22. A. N. Shniperov and K. A. Nikitina, "A Text Steganography Method Based on Markov Chains", *Automatic Control and Computer Sciences*, Vol. 50, No. 8, pp. 802–808, 2016.
23. M. Khairullah, "A Novel Steganography Method using Transliteration of Bengali Text", *Journal of King Saud University – Computer and Information Sciences*, In Press, 2018.
24. Binary convertor-http://www.asciitohex.com/
Mansi S. Subhedar, Vijay H. Mankar, "Image steganography using redundant discrete wavelet transform and QR factorization", *Computers and Electrical Engineering, ACM Digital Library*, (2016), Vol. 54, Issue C, pp. 406-422.
25. Su Q , Niu Y , Zou H , Zhao Y , Yao T . "A blind double color image watermarking algorithm based on QR decomposition", *In: Multimedia Tools Appl*, 72. pp. 987–1009.

AUTHORS PROFILE

Sanjive Tyagi, is perusing Ph.D. in Computer Application from Teerthanker Mahaveer University (TMU), Moradabad, India. He has done M.Tech. (CSE) in 2006, MCA in 2003 and M.Sc. Physics from CCS University, Meerut. He is working as an Associate Professor in Subharti Institute of Technology and Engineering (Swami Vivekanand Subharti University), Meerut. He has 16 years of teaching experience in various colleges. He has published various research papers as- 1. A High Capacity PDF Text Steganography Technique Based on Hashing Using Quadratic Probing in International Journal of Intelligent Engineering and Systems (IJIES- SCOPUS Indexed), 2019. 2. A Novel PDF Steganography Optimized Using Segmentation Technique in International Journal of Information Technology, Springer, 2019. 3. Secured High Capacity Steganography using Distribution Technique with Validity and Reliability- in IEEE Conference-SMART-2016, 4. High Capacity Steganography Protected using Shamir's threshold scheme and Permutation Framework", International Journal of Innovative Technology and Exploring Engineering (IJITEE- SCOPUS Indexed) and published 9 research paper on Digital Steganography in International Journal and 1 research paper on Smart Card Fraud Prevention Scheme in International Journal. He has attended various workshops, seminars and short terms courses in the area of computer science & engineering. His area of interest includes Information Hiding, Cryptography, Pattern Recognition, Algorithms Design and Data mining.

Dr. Rakesh Kumar Dwivedi, is Professor and Principal in College of Computing Sciences and Information Technology at Teerthanker Mahaveer University, Moradabad, India. He has completed his Ph.D. degree in the area of Digital Image Processing from Indian Institute of Technology Roorkee. He has done his M.Tech. degree in CSE from H.B.T.I. Kanpur, Uttar Pradesh. He has 22 years teaching experience in the field of computer science & engineering. He has published various research papers as 37 papers in International Journals of high impact factor and 38 research papers in Conferences in India and Abroad. He has chaired the session in the World Congress of Computer Science and Computer Engineering Conference, 2014 at Las Vegas, USA. He has participated in 19 AICTE approved short term courses in the area of Computer Science. His research area of interest includes the soft computing; Fuzzy based Hybrid Soft Classification, Parameter Optimization, Algorithm Design and Uncertainty Reduction using Fuzzy Techniques. Projects & Consultancy Projects includes In-house software development. His teaching & research areas includes Computer Algorithm, Java Programming, Database Management System, Soft Computing, Digital Image Processing and Soft Classification.

Dr. Ashendra Kumar Saxena, is Associate Professor in College of Computing Sciences and Information Technology at Teerthanker Mahaveer University, Moradabad, India. He has completed his Ph.D. from Mahatma Jyotiba Phule Rohilkhand University, Bareilly, India. He has done his master degree from Uttar Pradesh Technical University, Lucknow. He has 15 years teaching experience in the field of computer science & engineering. He has attended various workshops, seminars and short terms courses in the area of computer science & engineering. He has published various papers in national, international conferences and journals. His research area of interest includes the Optimization Techniques, Information Security, Operation Research, Algorithms Design and Cloud Computing.