IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol.-6, Issue-7, July 2018 E-ISSN: 2347-2693

Enhanced Image Transferring Scheme using Security Techniques

Madhura. M[1*], Mohana Kumar. S[2]

[1,2]Research Scholar, M.S.Ramaiah Institute of Technology, Bangalore, India

[*]*Corresponding Author: madhura.2901@gmail.com*

Available online at: www.ijcseonline.org

Accepted: 23/Jul/2018, Published: 31/July/2018

Abstract— Rapid growth in technology has enabled cyber attackers hack the data and misuse the content transferring using the authentication method. Due to this information transferring requires more security. Attackers can thoroughly figure out sensitive data from the traditional methods, using the hash value. Traditional methods of visual cryptography approach were used to transfer the sensitive data.

Enhanced image scheme is proposed to hide the sensitive data. Transferring sensitive data in the form of image with the printed text requires more security without allowing the attackers to change the sensitive data. In this project OCR engine, tesseract approaches help in recognizing and conversions of the printed text to the machine typed characters. AES algorithm is then applied to encrypt these machine typed characters. Steganography technique is used for binding the secure data with the cover image without changing the input data format. Digital signature approach is used for verifying the extracted and decrypted character from the output image with the input data. Using multiples algorithms in the proposed approach enables more security for transferring of sensitive data. PNG image format is used for the implementation of the proposed system for better accuracy.

Keywords— OCR engine, Tesseract, Cryptography, Steganography and Digital Signature.

I. INTRODUCTION

The rapid growth of the human being knowledge in understanding technology, the cyber attackers cracking the sensitive data has been increased. The information transferring requires more secure due to attack, the message or the password transfer requires more security. Since the content will be less it will be easier for the attackers in this case it is more important to provide the security for the content in each and every step of transferring the data.

Traditional methods of transferring the image which contains sensitive data have been recognized with numerous flaws. Rapid growth in technology has enabled cyber attackers hack the data and misuse the content transferring using the authentication method. Due to this information transferring requires more security. The sensitive data such has DOB, place, password, user id etc., requires more security. Attackers can thoroughly figure out sensitive data from the traditional methods, using the hash value. Traditional methods of visual cryptography approach were used to transfer the sensitive data.

In this paper explains the security can be provided for the image (sensitive data) which contains the printed text of message or password. This can be transferred from the sender to the receiver without allowing the attackers to identify the content. The content extraction from the image of the printed text of the message or the password by using Adaptive thresholding, OCR engine and Tesseract. These approaches help in recognizing the characters and conversions of the printed text of the machine typed characters. Later advanced encryption algorithm applied to secure the sensitive data.

This research paper is organised in the following manner. Section I contains the introduction of proposed topic, Section II contains related work of proposed topic and existing approach, Section III describe the methodology, Section IV describe the system implementation, Section V describe the algorithm, Section VI describe the result and evaluation of the proposed method, Section VII describe the conclusion with future works.

II. RELATED WORK

We present the brief summary of the earlier work carried out in the field of cryptography and security.

Akhil et.al [2] discussed the architecture of the tesseract tool and OCR tool transym for the character recognition. In his paper an attempt of comparison is made and drawn accuracy of both tools and he concluded tesseract is better and faster for character reorganization and transym may provide better accuracy than tesseract.

Pijush Chakraborty and Arnab Mallik et.al [3] discussed the tesseract one of the best open source optical character recognition tool for the character recognition from the scanned document or the printed copy. API jortho which is used for removing the errors in the text generated from the tesseract. Then also conversion of the corrected text to the six dot cell Braille format after conversion the text also placed in the refreshable Braille display. He concluded that reorganization accuracy increases in by converting colour image to gray scale image.

Chirag Patel, Atul Patel, Dharmendra Patel et.al [4], discussed the data hiding in the appropriate multimedia carrier known as the covers e.g. audio, video and image using steganography. Presented the background discussion of the major algorithm of digital image steganography. Image realization which is one of the most significant techniques of hiding information from the cover image without secret embedding.

Abdel-Karim Al Tamimi et.al [13], discussed the different encryption technique Symmetric key, asymmetric key and hashing key. He also evaluated the five different encryption algorithm AES, DES, RC4, BLOWFISH, RSA, also explained the characteristics of the different cryptography algorithm. Performance comparison of the different algorithm is well explained. His conclusion was that AES algorithm consumes less encryption and RSA algorithm consumes more encryption for decryption AES algorithm is enhanced than all other algorithm.

Avinash Kak et.al [6], discussed the AES encryption algorithm for the input data in the form of text and image. He has explained the all the steps detail with example.

Kundankumar Rameshwar Saraf et.al [8], discussed the Encryption and Decryption of the data in the form of text and image using Advanced Encryption standard. The author has used the code block chaining (CBC) approach with PKCS 5 padding for the image encryption. The text encryption been used is 128bit size of keys and plaintext. JPEG format of the image been used for the input data. He has successfully concluded the encryption of the text and image format.

José Manuel Ortega et.al [9], has explained the Cryptography, Hash function, Steganography and the digital signature technique in the python language with the detail description of the RSA and AES data encryption algorithm.

Pri a Bharti, Roopali Soni et.al [10], has explained the data hiding using the steganography and the cryptography. RSA algorithm is been used for the encryption of the data. LSB technique is used for embedding the information in the cover image. The combination of both the technique as satisfied the security, robustness and the capacity for the data transmission.

Li, X., Wang et.al [12], has explained the Particle swarm optimization (PSO) algorithm for improving the quality of the stego-images. The message is then hidden in the DC-to-middle frequency components of the cover-image. The comparison of the JPEG-based stenographic algorithm and the PSO with respect to the message capacity, better image quality and security level. According to the Experimental results PSO algorithm was better.

S.R. Subramanya and Byung K. Yi et.al [15], has explained the digital signature for the message, signature creation workflow with multiple signatures. He also explained generating the private and public key for the data transfer. RSA algorithm is been used for generating the public key and private key. He concluded that with encryption and decryption of message using the digital signature is the fast signing technology.

Laurent Luce's Blog et.al [16], has explained the Pycrypto a python package for the cryptography. SHA-256 hashing functions for the encryption of the data and comparison of the different hash functions. Digital signature for verifying the signed and unsigned data. The author has briefly as been explained all the main security techniques.

All the security systems provide the security using cryptography, visual cryptography, network security, steganography for the secure password. The password process using visual cryptography and OCR is the current system which will give you more secure than the cryptography. In case of visual cryptography technique provides the security for the user ID and password by using the Cryptographic technique i.e. visual cryptography. The registered user to the server inputs the password and ID to the device and original image will be created and save the image in the device. By adapting the visual cryptography the first shared image will be created with the pseudorandom generator along with the SEED containing the salt of ID and password. The user will send the ID and image to the server through security channel and abolish the image. Server will not be having idea about the password because it is difficult to retrieve password from one shared image.

For the successful login to the server the user needs to send the second shared image constructed using original image and the first shared image. The server overlaps both the first and second shared images and removes the background of the overlapped image and retrieves the ID using OCR. The server confirms both saved ID in the server and the extracted ID from the overlapped image and concludes success or fail,

result will be sent to the user. Figure 1 shows the data flow transaction using visual cryptography and OCR.

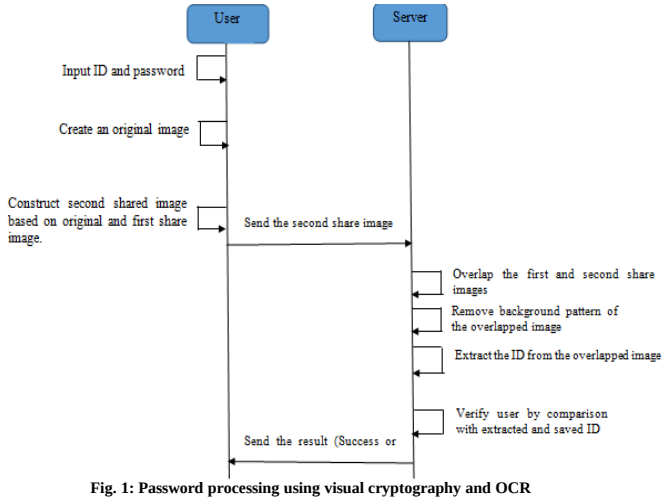


Fig. 1: Password processing using visual cryptography and OCR

III. METHODOLOGY

1. *Feature Extraction*

Firstly the printed texts of message or password need to be converted into the editable text. OCR which allow the scan printed text, or typewritten text to the editable text which can be reused or used in the other documents. Tesseract is one of the OCR engine, it is an open source and works best for the clean segmentation of the background and also the segmentation needs to be with high resolution. After the segmentation the input image cannot appear pixelated. For recognizing the text the tesseract will struggle after characters appears pixelated [2] [3] [4].

2. *Cryptography*

Symmetric key cryptography algorithm is been used for the securing the data, a same key will be used for the decryption and encryption of the data [5]. AES (Advanced Encryption Standard) is a symmetric key encryption is used for securing information [6][7]. Dispatcher uses encryption algorithm, recipient uses decryption algorithm. Encryption key and the encryption string will be generated using the AES algorithm. Encryption key will be shared to the recipient for the decryption of the data [8].

3. *Steganography*

It is a technique of hiding the data behind the image audio, image, video etc. Apart from the authorized send and the receiver no one will be aware of the hidden data [9]. Image steganography is one of the technique in which message will be hidden behind the image [10]. To make more secure of the message encryption technique will be used so the hidden text is also unreadable, but still it is exits as data [11] [12]. LSB (Least significant bit) based algorithm is used for hiding the message by embedding it in a cover media like images which modify the pixels of the image. LSB substitution can be used for both the greyscale and coloured image.

4. *Digital Signature*

It is a technique used to validate the integrity and authenticity of the message, document and software. It is digital equivalent to the handwritten signature or stamped seal. It provides the assurance of evidence to original data [14]. Digital Signature is based on the asymmetric cryptography, uses RSA (Rivest Shamir Adelman) algorithm. First the private key and the public key will be generated for the hash code. Then using the private key the string is signed which is generated from the image. The public key will be shared with the recipient but not the public key. The signed string will be in the byte format that will be converted to the hexadecimal format using hexifying.

To validate the decrypted string public key will be used to validate the signed text and the decrypted string.

IV. SYSTEM IMPLEMENTATION

Many people will use the short length password or same user id or knowingly user id or user names in multiple system or neglectful password in multiple systems this causes consequently cyber accidents which occurred often cause less security, hence require more traditional password conversation scheme which enhanced password processing scheme based on visual cryptography and OCR.

We suggest enhanced password processing scheme based on image using cryptography different traditional approach uses hash value and text value. Our proposed retires the image contents and encrypts the sensitive data and appending the encrypted string into the other image. Decryption of the append image for separating the encrypted string from image. Decryption of the encrypted string.

Data processing is as follows:

1. The user inputs image which contains the data.
2. Adaptive thresholding is applied to convert input image to grey scale image, if the input image is coloured image and to remove the noise of the image.
3. OCR engine tesseract will recognize the characters from the image and conversion of the printed text to the machine typed text.
4. The text is signed using Digital signature for the verification of the data. Public key will be shared to the server for the decryption process.

5. Symmetric key cryptography generates the encryption key and the string for the machine typed characters using Hash codes. Encryption key will be shared to the server for the decryption process.
6. Encrypted string is concealing to the other image using steganography so the data ready to send for the server will be in the input data format.
7. Decryption is done in the reverse order from Steganography, cryptography using shared encrypted key.
8. To validate the decrypted string public key will be used to validate the signed text and the decrypted string.

Figure 2 shows the data flow from the client side and figure 3 shows the data flow from the server side.

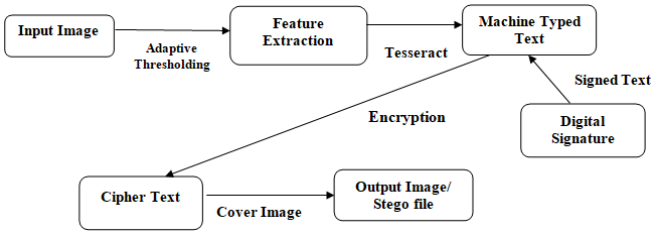


Fig. 2: Process in the Client Side (Encryption of the Data)

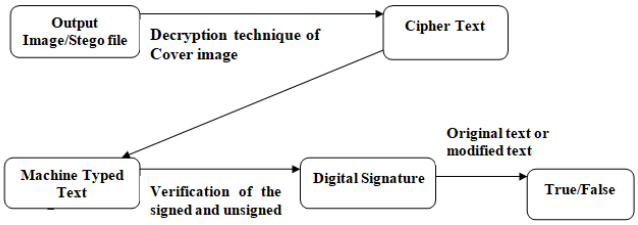


Fig. 3: Process in the Server Side (Decryption of the Output Image)

V. ALGORITHM

1. *RSA Algorithm for the encryption and decryption*

- Take two different, large primes p and q Preferably these have a similar byte-length
- Multiply p and q and store the result in n
- Find the totient for n using the formula
  $\Phi(n) = (p-1)(q-1)$
- Take an e co prime that is greater than 1 and less than n
- Find d using the formula
  $d \cdot e \equiv 1 \bmod \varphi(n)$

At this point, the pair (e, n) is the public key and the private key (d, n) is the private key.

2. *Digital Signature for signing the message using python*

```
From Crypto. Hash import SHA256
From Crypto.PublicKey import RSA
From Crypto import Random
Key = RSA.generate (1024, random generator)
Hash = SHA256.new (text).digest ()
Signature = key. Sign (hash, '')
```

VI. RESULTS AND EVALUATION

The input image in the PNG format, the image can be the screenshot image, scanned image, captured image or the cropped image but the format of the image should be PNG format. As in the PNG the whole image will not be considered as the object and easily we can identify and extract the characters or the image. In the other image format like JPES, TIFF etc the whole image will be considered as the image and clarity of identify the characters will be difficult. For the better accuracy of the character extraction in the OCR tesseract along the number of characters extraction from the input image with 100% printed grayscale image is considered for the implementation in the proposed model.

Fig. 4: Screenshot image

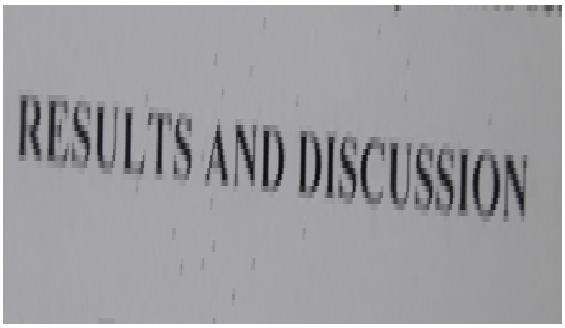


Fig. 5: Captured image

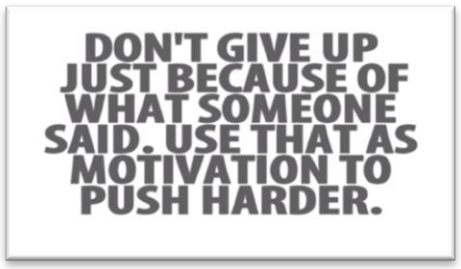
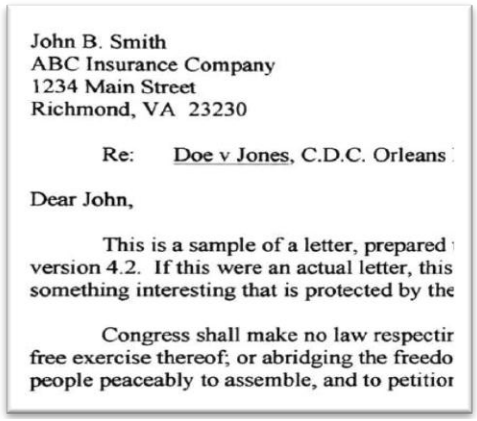


Fig. 6: Downloaded image


John B. Smith
ABC Insurance Company
1234 Main Street
Richmond, VA 23230

Re: Doe v Jones, C.D.C. Orleans

Dear John,

This is a sample of a letter, prepared
version 4.2. If this were an actual letter, this
something interesting that is protected by the

Congress shall make no law respectir
free exercise thereof; or abridging the freedo
people peaceably to assemble, and to petitior

Fig. 7: Scanned image

The Figure 4 is the screenshot image in the PNG format, figure 5 is the captured image in the JPEG format and full noised image, figure 6 is the downloaded image in the PNG format and figure 7 is the scanned image. If the input image is PNG format characters can be easily and clearly recognized. Digital Signature is applied for the extracted characters using RSA algorithm, public key will be sent to the server side for the decryption process.

Lenna.jpg Hiddendatalenna.jpg

Fig. 8: Cover images and Output images

The encryption technique is applied for the extracted character using AES algorithm then the encrypted string is bind with other output image. Figure 8 contains the cover image and output image jpg format cover image of size 445 KB and output image of size 447 KB with encrypted string binded to the cover image. The output image can be in any of the format once the string is bind to the image then the size of the image will vary. Variation of the output image size before and after binding of the encrypted string is shown in the table2. Mean variation and the deviation of the cover image and the stegno image is shown in the table 2. Encryption string of the test input image is

""c58c23e9aad108e423e2ad0ccb261c7563e03fb9a6 a3 1217aa25c2cb905640d7"

Table 1: Output image size before and after bind of the string

Image Format	**Cover image size**	**Encrypted data format**	**Bind image size**	**Variation of the image size after data is bind**	**Time taken for the image to bind**
JPEG	80.7 KB	String	87.89 KB	7.19 KB	Normal
PNG	10.0 KB	String	25.76 KB	15.76 KB	Normal
JFIF	254.9 KB	String	10.47 KB	244.43 KB	Normal
TIFF	321 KB	String	421.09 KB	100.09 KB	Normal
RAW	175 KB	String	537.4 KB	362.07 KB	Normal

The Output image is decrypted by separating the cover image and the Cipher text. The encrypted string will be decrypted and will be converted into the original extracted character. For the verification of the decrypted character. The proposed model is implemented on spider IDE, Anaconda distribution using Pyton Language.

VII. CONCLUSION AND FUTURE WORK

In this paper we have presented cryptography, steganography and digital signature for providing the security for the data present in the input image. While transferring the image content from client to server more security need to be

provided for the sensitive data to secure from the attackers. Hence we have used cryptography technique to encrypt the image content (sensitive data), Steganography technique used to bind the encrypted string to output image for the purpose of input data format and digital signature for verifying the sensitive data after decryption. Performance of the JPG image format is decrypted with full accuracy compared to other image format.

In Future work can be implemented by using the different image formats as in our paper we have implemented using the JPEG format. Counter marking of gray scaled Images that are suitable as Cover for the future data hiding scheme. Investigate work on AVI Steganography based on proposed scheme is in process.

REFERENCES

[1] Dana Yang, Inshil Doh, and Kijoon Chae "Enhanced Password Processing Scheme Based on Visual Cryptography and OCR" Dept. Computer Science and Engineering Ewha Womans University, 2017,IEEE.

[2] Akhil, "Overview of Tesseract OCR engine" Department of Computer Science and Engineering National Institute of Technology, Calicut Monsoon-2016 Seminar Report.

[3] Pijush Chakraborty and Arnab Mallik "An Open Source Tesseract based Tool for Extracting Text from Images with Application in Braille Translation for the Visually Impaired" International Journal of Computer Applications (0975 –8887) Volume 68 – No.16, April 2013.

[4] Arnab Mallik, Asst. Professor, CSE Dept, Calcutta Institute of Engineering and Management "An Open Source Tesseract based Tool for Extracting Text from Images with Application in Braille Translation for the Visually Impaired", International Journal of Computer Applications (09758887) Volume 68–No.16, April 2013.

[5] Chirag Patel, Atul Patel, Dharmendra Patel, Charotar University of Science and Technology "Optical Character Recognition by Open Source OCR Tool Tesseract: A Case Study" international Journal of Computer Applications (0975 –8887) Volume 55–No.10, October 2012.

[6] Md. Alam Hossain, Md. Biddut Hossain, Md. Shafin Uddin, Shariar Md. Imtiaz "Performance Analysis of Different Cryptography Algorithms" Computer Science and Engineering Department, Jessore University of Science & Technology, Bangladesh.

[7] Kundankumar Rameshwar Saraf, "Text and Image Encryption Decryption Using Advanced Encryption Standard" International Journal of Emerging Trends & Technology in Computer Science (IJETTCS).

[8] José Manuel Ortega, "Python Cryptography & Security" europython 2016, bilbao, 20-26 July.

[9] Pri a Bharti, Roopali Soni, "A New Approach of Data Hiding in Images using Cryptography and Steganography", International Journal of Computer Applications (0975 – 8887) Volume 58–No.18, November 2012.

[10] Chirag Patel, Atul Patel, Dharmendra Patel "Optical Character Recognition by Open Source OCR Tool Tesseract: A Case Study" International Journal of Computer Applications (0975 – 8887) Volume 55–No.10, October 2012.

[11] Li, X., Wang, J.: A Steganographic Method based Upon JPEG and Particle Swarm Optimization Algorithm. Information Sciences 177(15), 3099–3109 (2007).

[12] Abdel-Karim Al Tamimi, "Performance Analysis of Data Encryption Algorithms" International Journal of Advanced Research in Computer Science and Software Engineering, Volume 6, Issue 3, March 2016.

[13] Avinash Kak, "AES: The Advanced Encryption Standard Lecture Notes on "Computer and Network Security", Purdue University, February 2018.

[14] S.R. Subramanya and Byung K. Yi, "Digital signatures" March/April 2006 IEEE Potentials.

[15] Laurent Luce's Blog "Python and cryptography with pycrypto" March 2018.

[16] Mohana kumar S and Jagadeesh S N "Study of privacy preserving and detection of sensitive data explore using message digest" International journal of Advance research in computer and communication, volume 5, issues 6, pages 118-122.

[17] N.S.Lele,"Image Classification Using Convolutional Neural Network", International Journal of Scientific Research in Computer Science and Engineering, Vol.6, Issue.3, pp.22-26 , 2018.

Authors Profile

Madhura M, She received B.E degree in Computer Science and engineering from Visvesvaraya Technological University in the year 2016. M.Tech in Software engineering at M.S.Ramaiah Institute of Technology, Bangalore in the year 2018. Her research interest is in Image Processing, Computer Vision, Machine Learning.

Mohana Kumar graduated from Visvesvaraya Techonology University(VTU) and Ph.d from VTU in the year 2018. His current working as Associate Professor in M.S.Ramaiah institute of techonology. His areas of interest Protocol engineering Network security and software patterns.