Improved Secret Information Hiding using SHA-256 and Invisible ASCII Character Replacement Technology

Swati C. Dandekar
Department of Computer Engineering
DYPIT,Pimpri
Pune, India
swaticdandekar@gmail.com

Prof. Prashant G. Ahire
Department of Computer Engineering
DYPIT,Pimpri
Pune, India
ksprashantahire@gmail.com

Dr. Jyoti Rao
Department of Computer Engineering
DYPIT,Pimpri
Pune, India
jyoti.aswale@gmail.com

***Abstract*—This is the information era . Lots amount of information is get transmitted from one end to another. The text document exchange is most widely used technique over internet for communication. This may contain some valuable information or some general data. If these text documents contain valuable information and if it gets hacked by some unauthorized user then they may change it or make misuse of this information. So that we need some technique to hide our important and valuable text information while communicating over internet. Still there are many hiding technique present also proposed by many authors but these having some limitation and drawbacks like poor robustness, lower embedding rate and semantic clutter. Due to all these weakness in existing system, interceptors may hacked or extracted the information. To overcome all these drawbacks of existing system, we proposed a new technique based on the invisible ASCII character replacement method the hash function. In this approach we first find the binary formatted information , after that the space character in every carrier is replace by SOH with the help of some replacement algorithm. In third stage a hash value is generated. At last the hash value is compare with the encoded secrete information**

Keywords—Information security, Invisible character replacement technology, Hash function, Steganography, Textbased information hiding.

I. INTRODUCTION

Day by day the use of internet increasing rapidly, so it become reliable and the most efficient way of communicat-ion. The text-document exchange is Most widely used technique over the internet for communication. While transmitting text valuable data over a public network during communication, this technique has some challenges. There may be possibility of information hacking by interceptor hence there is a need to safely transmission of this informa-tion over public network. Many techniques and algorithms are proposed like MAC, SHA, RSA, AES and so on, to provide reliable and secure communication over the internet through text documents. The data encrypted by using these methods has the code Complexity and these techniques uses some format while encryption, so it helps eavesdropper to find the correct one decryption technique and extract information from it. We proposed a novel text-based information hiding technique called secrete information hiding techniques based on invisible ASCII character replacement and hash function, to overcome all drawbacks of current techniques. The information hiding method basically classified into two types, one is digital watermarking and second is steganography. For information hiding purpose in our approach, we consider only steganography. Steganography is data hidden within data. Steganography techniques can be applied to images, a video file or an audio file, emails. These technique makes use of carrier document, that carrier may be text file, message, video, audio. Hence to hide secret information, steganography use any one carrier to transmit the secret information within this carrier files. Based upon the use of carrier document steganography is classified into two types, one is multimedia based steganography and second is text based steganography. In multimedia steganography lots of space available, this technique is robust and very useful for the data transmission over the internet. In text-based steganography less space, low embedding rate and easy identification of the secret information happen, so very little work has been done on the text-based steganography technique. But the text document is most widely used for communication, there are many challenges to improve the text-based steganography technique. To overcome the disadvantages of text-based steganography, we have considered text-based steganography as a study.

A text-based information hiding method is classified into two types, one is format-based information hiding method and second is content based information hiding. In a format based text hiding method the secret information is hide by adjusting the font size, line space, word space, word count and so on. This method is very rarely used because changing format of carrier document will affect secret information. Second method is the content-based information hiding method. This method is completely based on syntax, semantic and statistical properties of natural language.

In our approach, we first convert secret information into binary format then, after that with the help of some replacement algorithm the space character in every carrier is replaced by SOH. In the third stage, a hash value is generated. Then at last the hash value is compared with the encoded secret information.

978-1-5386-5257-2/18/$31.00 ©2018 IEEE

The rest of paper is organizes as follows. Section II introduces review of literature. Section III system architecture of proposed system. Section IV system overview. We conclude the paper in Section V.

II. REVIEW OF LITERATURE

Feng Liu1, Pengpeng Luo, Zhujuan Ma, Cheng Zhang , Yiwen Zhang in[1] proposed the basic concept of information hiding based on hash function and ASCII character replacement technology. They also explained the drawbacks of some of existing techniques. Here they use SP and SOH character to hide secret information. Space character SP in carrier segment are replaced with ”SOH” by replacement algorithm.

H. Krawczyk, M. Bellare, and R. Canetti in [2] proposed a technique Keyed-Hashing for Message Authentication that is HMAC. In this approach, they have described use of HMAC with cryptographic hash function such ad MD-5 and SHA-1. In order to legalize the transmission of information between two parties, a message is sent between these parties that share a secret type of information . The length of key used in HMAC is not fixed it vary .The security of the message Authentication mechanism is depends on cryptographic properties of the hash function H.

FEDERAL INFORMATION PROCESSING STAND - ARDS PUBLICATION[3] Specifies a new function SHA-3 that is Secure Hash Algorithm 3. Hash function generate output value of fixed length. Basically SHA-3 consist of four cryptographic hash function two extendable-output functions. cryptographic hash functions are SHA3-224, SHA3-256, SHA3-384, and SHA3-512and they generate fixed length digest such as 224, 256, 384, 512bits respectively .

ADVANCED ENCRYPTION STANDARD stand for AES given by Federal Information Processing Standards Publications (FIPS PUBS)[4] are issued by the National Institute of Standards and Technology. AES is symmetric encryption algorithm. In AES the length of keys used are 128 , 192 , 256 bits. In AES algorithm the basic unit for processing used is a byte, a sequence of eight bits treated as a single entity. Arrays of bytes is used in AES . The input, output and Cipher Key bit series are developed using this arrays of bytes.

R.L. Rivest, A. Shamir, and L. Adleman[5] proposed a method for obtaining public key cryptosystem and digital signature. The public-key cryptosystem ,is based on the concepts given by Diffie and Hellman. The public-key cryptosystem must be implemented with trap-door one-way permutations to implement signatures, hence the decryption algorithm will be applied to unenciphered messages.

Pierre Moulin and Joseph A. OSullivan gives a theoretical presentation of information hiding techniques in [6]. Here they formalized some points also evaluate the hiding capacity which related to reliable transmission of secretes information. They specifies the basic rules for the quantities, that include information hiding rates, acceptable distortion level for information hider and attacker.

Mehran Andalibi and Damon M. Chandler gives a Digital Image Watermarking via Adaptive Log Texturization technique[7]. Here, they present a new algorithm for invisiblegrayscale logo watermarking. This algorithm operates on the basis of texturization of logo. In this they first do separation of the crowd image into poor and good texture parts. After that, using Arnold transform the good texture region is transformed into a similar texture and poorly texture is rotated through the lossless rotation. And at the end, each region is embedded via standard wavelet-based embedding scheme.

A Hide and Seek method is introduced by NIELS PROVOS AND PETER HONEYMAN in this paper[8] . Here they make discussion on the existing steganographic systems and there recent research in detecting them using numerical steganalysis. While other survey focus on the use of information hiding and overview on detection algorithm. They also discuss statistical steganalysis. For steganography they only consider JPEG image.

An algorithm called Text Information Hiding based on replacement technique is presented by Gongshen Liu, Xiaoyun Ding, Bo Su, Meng Kui[9]. The text based informati on hiding is not that much easy because text file does not contain any redundant information . Here they propose an information hiding method based on the substituted conception. They told that any word in text information is replaced with any meaning full similar conception but keeping the meaning of the sentence same as previous.

A.A. Mohamed gives an improved algorithm for information hiding based on features of Arabic text: A Unicode approach[10]. Their proposed algorithm supports a high capacity of the carrier media also the embedding capacity rate ratio of this algorithm is high .In Arabic, there is a group of six letters written isolated in a text. An isolated letters used by them as hidden key in Arabic text written in Unicode format. Such a letters are searched by the replacement techniques for the replacement in carrier text document.

III. SYSTEM ARCHITECTURE

The architecture of information hiding system is shown in figure. This architecture consist of two modules first is sender module and another one is receiver module.

Basically there are again two different modules in sender module one is information hiding module and another is encoding module . For the secrete information hiding process here we are using a plane English text document. At the encoding module division of a binary secrete message take place. This text document and encoded secrete message is then provided to information hiding module. In information hiding module the hidden characters present in text document is replace by the encoded secrete information.

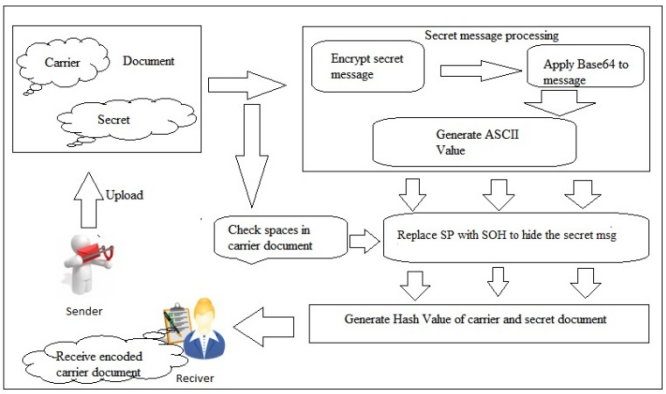

Fig 1: Architecture of Information Hiding System

After performing hiding process, text document is sent over public network like internet. To get the secrete message exact reverse operation is done at the receiver module on received text document. First of all extraction process is done on text document. After that decoding process is performed on extracted secrete information which results in original secrete information.

A. Mathematical Model

System can be represented as :

SA= {IT, IH, ITS, IE, CD, SI, EMB, S_{key} , Sender, Receiver , EP, SIE, $E_{Procedure}$}

Where,

SA=Start Application

$$\text{EMB}=\int_0^n \text{CD} \in \text{CD}_{\text{segmentation}} \quad (1)$$

$$(\text{Sender} + \text{Receiver})\ S_{\text{key}} \quad (2)$$

$$\text{EP}=\int \sum_{k=0}^{n} \text{SIE} + \text{IS} + \text{RSH} \quad (3)$$

$$\text{E}_{\text{Procedure}} \approx ! \text{ EP} \quad (4)$$

IT=Information Transmission

IH=Information Hiding

ITS=Information Transmission Stage

IE=Information Extraction Stage

$$\text{CD} \in \text{General Information} \quad (5)$$

CD=Carrier Documents

SI=Secrete Information

EMB=Embedding Procedure

S_{key}=Shared Secrete Key

EP=Encoding Procedure

SIE=Secrete Information Encoding

IS=information Substitution

RSH=Replaced Segment Hashing

$E_{Procedure} = Extraction\ Procedure$

VI. SYSTEM OVERVIEW

The existing encryption algorithm has some drawback like code complexity, generates messy code. These techniques follow some format while encryption, so it comes easy for the eavesdropper to find the correct decryption technique and extract information from it. Also there are some issue in content based information hiding algorithm. The natural language processing that is content based information hiding is realized by processing the syntax, statistical property of natural language. It makes use of substitution table, but the use of this table is not secure. Since the substitution table can be hacked by interceptor. In the proposed system, we have focused on the text information hiding based on hash function ASCII character replacement. To improve level of security here we are using SHA-256 instead of MD5. In proposed work first upload secret information and carrier text then apply hash function to carrier document & to our secret msg. Next apply encryption algorithm to secret information and generate key for that information then to increase the level of security we are applying Base64 algorithm to encrypted secret information, after that convert that encrypted value into ASCII value. Then convert it to binary format. Next with the help of some replacement algorithm the space character in every carrier is replaced by SOH. Finally sent that document to receiver on public network. Receiver just needs a reverse process to extract secret information.

A. Methodology of Evaluation

We use Java and MySQL as the database for the implementation and run on 2.33GHz Intel Core 2 Duo Processor machine with 2 GB RAM. Eclipse IDE for the Development of the proposed approach. Apache Tomcat Server for target runtime as the local host server.

B. Metrics of Evaluation

MER represent the Maximum Embedding Rate means the ratio of the theoretical maximum secret bits that could be embedded in one carrier document. The ratio is influenced by the quantity of the secret message length and no of spaces in carrier document. Here we have considered a complete document as one segment. Because of this the length of secret message in our experiment is not fixed but the number of bit to be replaced in one document is fixed. Therefore according to the definition of MER, MER for this experiment is calculated as

$$\text{L}_\text{m} = \text{ASCII(base64)} * 8 \quad (6)$$

$$\text{MER} = \frac{\text{Lm}}{\text{SP in carrier document}} \quad (7)$$

Where ,

Lm represent the length of secret message

base 64 represent the encrypted message

TABLE I. MER OF OUR METHOD

Lm	0-15	16-31	32-47
Base 64	32	64	96
Required SP	256	512	768

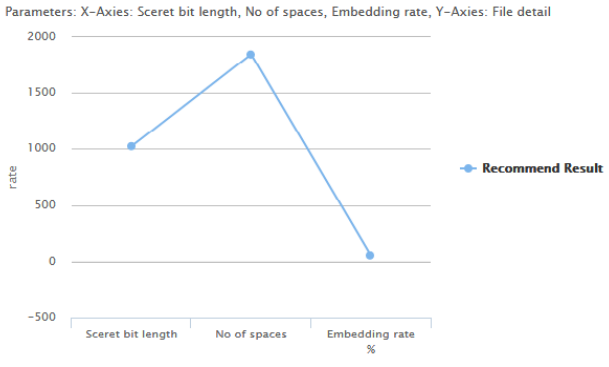

Fig 2: Message embedding rate

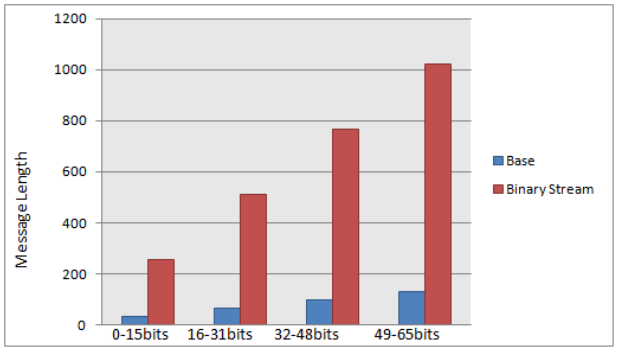

Fig 3: Expected SP for embedding msg into carrier document

Fig2 depict the message embedding rate which is calculated by using formula no 7. Fig 3 depict expected SP for embedding secret msg into carrier document that is , it is the ratio of base to binary stream. Base is the value which is obtained by after applying BASE 64 algorithm to the encypted msg . Binary stream is the final binary output generated by application of all the algorithms.

V. CONCLUSION

Based on the analysis of the text-based information hiding technology, in this paper we proposed a improved approach which secure text data. Here we are using text base hiding method based on the combination of invisible ASCII character replacement technology and SHA 256 .

ACKNOWLEDGMENT

Success is never achieved Single-handed. I am heartily thankful to my guide Prof. Prashant G. Ahire for his valuable guidance and inspiration. In spite of his busy schedules he devoted his self and took keen and personal interest in giving me constant encouragement and timely suggestion. His guidance always helps me to succeed in this work. I am also very grateful for his comments while designing part of my research paper and learned many things under his leadership.

References

[1] Swati Dandekar and Dr.JyotiRao,"A Survey on Improved Secret Information Hiding using SHA-256 and Invisible ASCII Character Replacement Technology", International Journal Of Advance Research, Ideas And Innovations InTechnology, Vol.4,Issue-1,January2018.

[2] Feng Liu1, Pengpeng Luo1, Zhujuan Ma2, Cheng Zhang1, Yiwen Zhang1, Erzhou Zhu1,Security Secret Information Hiding Based on Hash Function and Invisible ASCII Characters Replacement. 2016 IEEE TrustCom / BigDataSE / ISPA 2324-9013/16

[3] Hugo Krawczyk, Mihir Bellare, Ran Canetti. HMAC: Keyed-Hashing for Message Authentication. RFC 2104, February 1997.

[4] SHA-3Standard: Permutation-Based Hash and Extendable-Output Functions, FIPS PUB 202. National Institute of Standards and Technology (NIST). August 2015.

[5] AES-The official Advanced Encryption Standard, FIPS PUB 197. Computer Security Resource Center. National Institute of Standards and Technology (NIST). Retrieved 26 March 2015

[6] Ronald L.Rivest, Adi Shamir, Leonard M.Adleman. A Method for Obtaining Digital Signatures and Public-Key Cryptosystems. Communications of the ACM (Association for Computing Machinery) 26(1), 1983, pp. 9699.

[7] Pierre Moulin, Joseph A. O'Sullivan. Information-theoretic analysis of information hiding. IEEE Transactions on Information Theory, 49(3), 2003 pp563-593.

[8] Mehran Andalibi, Damon M.Chandler. Digital Image Watermarking via Adaptive Logo Texturization. IEEE Transactions on Image Processing, 24(12), 2015, pp.1-15.

[9] Niels Provos, Peter Honeyman. Hide and seek: an introduction to steganography. IEEE Security Privacy, 1(3), 2003, pp.3244.

[10] Gongshen Liu, Xiaoyun Ding, Bo Su, Meng Kui. A Text Information Hiding Algorithm Based on Alternatives. Journal of Software, 8(8), 2013,pp.2072-2079.

[11] AA Mohamed. An improved algorithm for information hiding based on features of Arabic text: A Unicode approach. Egyptian Informatics Journa,15(2), 2014, pp.79-87.