International Journal of Technical Innovation in Modern Engineering & Science (IJTIMES)
Impact Factor: 5.22 (SJIF-2017), e-ISSN: 2455-2585
Volume 4, Issue 7, July-2018

Security Secret Information Hiding Based on Bit Parity and Invisible ASCII Characters Replacement

Snehal Vilas Gadhave[1], Vikas Nandgaonkar [2]
[1,2] *Nutan Maharashtra Institute of Technology, Talegaon Dabhade, Pune, snehagadhave98@gmail.com*

Abstract— The frequently usage of the text document in the communication make the text-based information hiding technology still crucial topic in computer security. Currently, the popularly used techniques in this area are facing many problems, such as poor robustness, semantic clutter and lower embedding rate. As a result, the secret information can be easily detected or even extracted by the interceptors. To deal with these problems, this paper proposes a novel secret information hiding algorithm based on the integration of the hash function and the invisible ASCII character replacement technology. Firstly, the binary secret information is encoded with even number of "1" in each group. Secondly, the space characters in each carrier segment are replaced with "SOH" by the replacement algorithm. Thirdly, the replaced segment is processed by the hash function. Finally, the algorithm is completed by comparing the generated hash values with the encoded secret information. Furthermore, by utilizing the hash collisions of the previous segment, the algorithm is improved to enhance the embedding rate and security. The experimental results demonstrated that the proposed algorithm is effective, feasible and reliable.

Keywords— Text-based information hiding; Invisible character replacement; Hashfunction; Information security.

I. INTRODUCTION

With the rapid development of the networking technology, Internet becomes the important and efficient channel to communicate for commons. However, there is a big challenge when the end users transmit their private information through public network, since this information is likely to be intercepted by the eavesdropper. So, the safety transmission of private information on the internet attracts much attention from both research communities and industry. In order to make secure communication, many encryption techniques have been proposed to protect the private data, such as MACs [1], SHA-3[2], AES [3], RSA [4] and so on. But the data, which is readable by human, would become a pile of messy code after encryption process. As it may express the existence of secret information, interceptors are likely to pay attentions on such frequently appeared encrypted code. In order to deal with this shortage, the technology of replacement, called information hiding [5], is proposed to ensure the transmission of secure information. Mainly, there are two kinds of the information hiding technology, the digital watermarking [6] and the steganography [7]. In this paper, we focus only on steganography. Steganography is the practice of concealing a message, file, video or image within another message, file, video or image. In this technology, many kinds of messages i.e. mail, video, audio, text & digital image can be used as carriers for secret information transmission. According to the carriers, the steganography technology can be classified into two below categories: the multimedia based (like video and digital image) steganography and the text-based (like .doc, .txt, and .pdf files) steganography. Meriting from the relatively lager redundant space in the carriers of multimedia, it is relatively easier for the former one to embed secret information. In addition, the multimedia-based steganography not only provides high rate of secret information embedding but also makes the embedded secret information which will be hard to detect. Very opposed in nature, due to smaller redundant space for secret information, relatively lower embedding rate and easier to figure out the change on the carrier documents, little work has been performed on the text-based steganography technology.

However, due to the frequent & wide usage of text document in the communication, the text-based steganography technology has still attracted much attention from the hiding of information from research communities. So, the way utilizing of text as the carrier to embed the secret information even so has its merits. According to the embedding style of secret information, the text-based information hiding mechanisms can be divided into two types: the format-based information hiding mechanism and the content-based information hiding mechanism. Notably, the format-based information hiding mechanism can embed the secret information to the carrier document with adjusting its font, word space, line space, words count in one line, adding the blank characters and so on. Even so, this method is relatively poor robustness and been rarely used in practice since the change of the format of the carrier document will directly lead to the secret information's disappearance.

The content-based information hiding mechanism, is also called natural language based information hiding. This is realized by processing the syntax (such as TEXTO algorithm, NICETEXT algorithm and synonymous replacement algorithm [8]), semantic (such as machine translation based algorithm [9]) and statistical properties (such as Markov Chain based algorithm [10]) of natural languages. Since the underlying natural language processing technology is far from mature by now, there are still some road blockers to be resolved in this technology. On another hand, by the present language processing algorithms, there is obvious distinction in between the original nature language and the generated carrier document. This difference can be easily identified by unaided eyes. So, the documents that generated by the present language processing algorithms can't meet the practical application requirement from the point views of syntax, statistical properties and semantic respectively. On the other hand, the complex features of natural language make it difficult work for constructing an effective and reasonable substitution table for the replacement based hiding algorithm. Even though an ideal substitution table is available, this algorithm is still not secure enough, since the substitution table can be hacked by interceptors with very little effort's. As well as most of the replacement-based hiding algorithms are accomplished by a simple way, such as "0" represents replacement and "1" represents unchanged or reverse. It is hard for this simple mode to resist attacks from detecting algorithms which is based on statistical analysis. To deal with issues existing in the current context-based information hiding algorithms, this paper proposes a new algorithm of secret information hiding by integrating the hash function and the invisible characters replacement technology. In this algorithm, firstly the secret binary information encoded for parity which used to distinguish secret-carrying sections with non-secret carrying ones and in the meantime enhanced the security of algorithm. Then the space characters in the English text segmentation are replaced with an invisible character called "SOH", one at a time by corresponding replacement algorithm. Then hash function is employed to compare the hashed result and secret information to complete the embedding procedure. The receiver just needs a reverse process to extract the secret information. The remainder of this paper is organized as follows. Section II briefly states the invisible ASCII characters. Section III tells about the Security challenges and Architecture. Section IV gives the idea about the methodology. Section V Implementation Details and Graphs. Finally, Section VI briefly concludes with the paper and outlines of our future work.

I. ANALYSIS OF ASCII CHARACTERS

ASCII is a computer coding system which is based on the Latin alphabet, basically for the display of modern English and other western European languages. Now a day, it is the most common single-byte encoding system, equivalent to the international standard ISO/IEC 646. At present, algorithm of information hiding based on the substitution of invisible characters are primarily focusing on adding spaces or line breaks to some specific locations [11] or replacing blank space by null character (code as 0000000) according to the secret information. For example, the famous information hiding system WbStego in the market is built based on these methods. But those methods are rarely used because of the poor robustness and relatively lower embedding rate of it. After performing many different tests, we could find that SOH (start of head, coded as 0000001) and SP (space character, coded as 0100000) have similar effects in most of documents. Simultaneously, there are a lot of candidate SPs in English document for replacing. So, in this paper, we use SP and SOH substitution method in English text to implement our information hiding algorithm.

II. SECURITY CHALLENGES

As most of the time data sent over the internet is in text format. Mostly this data contains some valuable data. If such file or data get hacked by hackers or some unauthorized person, that information may change by hacker or used in wrong way.

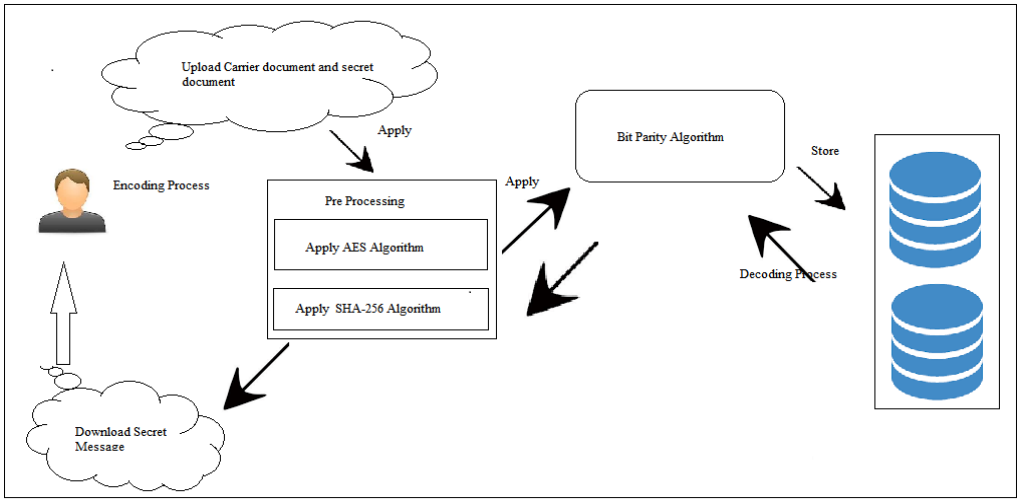


Fig 1. The workflow of the secret information transmission

Therefore, it is necessity to provide the security for the data which is send over the internet. There are many techniques exists currently like AES, DES, MD5, RSA which are used to hide secrete information in text document but have some drawbacks such as code complexity and these techniques follow some format while encryption so it becomes easy for the eavesdropper to find the correct decryption technique and extract information from it. Fig.1 outlines the overall secret information transmission workflow by applying our hiding algorithm. Generally, the workflow can be divided into 3 stages, the information hiding stage, the information transporting stage and the secret information extracting stage.

III. METHODOLOGY

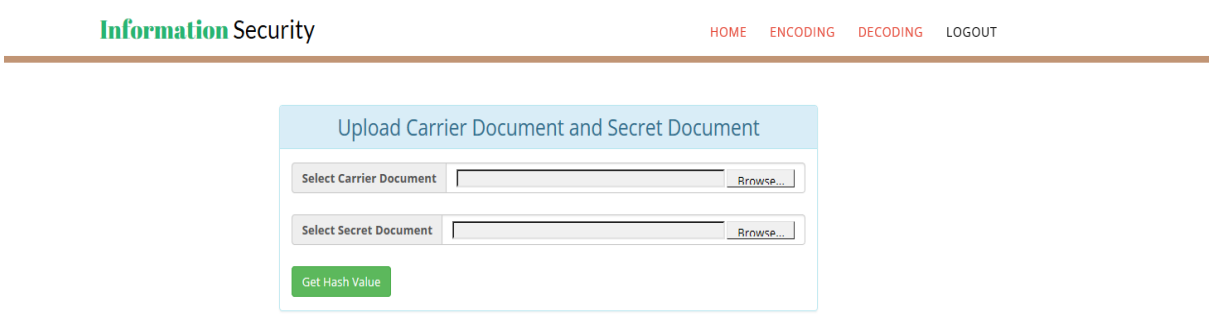
To beat out all the challenges, we are replacing MD5 to SHA-256 to send secret information successfully without getting hacked we proposed a new approach called ASCII character replacement technique for data hiding in a text document. In this accuracy will be increased with the help of parity algorithm. Here, firstly we find the segmentation of text document and add a parity bit to binary converted secrete information. After the even parity done on secrete information, a secret information hiding process is done on the carrier document of text. In this paper hiding process, each encoded binary secrete information is replaced by the ASCII character of the carrier text which is done by some replacement algorithm. Then a hash value is generated. At last the hash value is compared with the encoded secret information.

I. Sharing Settings

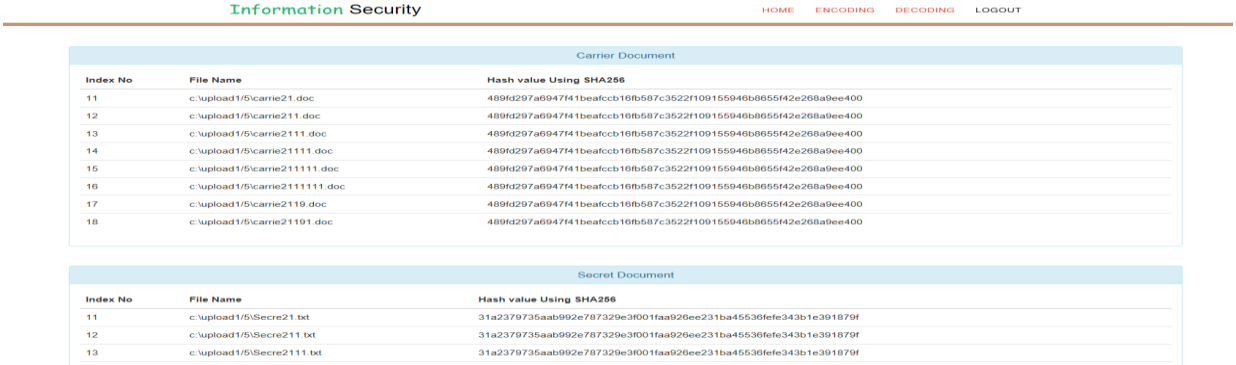
Before the secret message has been transmitted, some shared information, such as the division of carrier document and the division of the secret information, between the sender and receiver should be set. Furthermore, the sender and receiver also need to reach an agreement on the form of hash function. Some famous hash functions, such as MD4, MD5, and SHA-1, and even user defined hash functions are all feasible. In this paper, SHA-256 is selected as our hash function. Then AES is used to encrypt the secret information and Bit Parity Algorithm is used to generated secrete key for the hash value & secrete key generated by AES algorithm. While decoding, the data is encrypted by Bit Parity Algorithm. Once the decoding process starts Bit Parity Algorithm will be applied for the decoding the encrypted data. The decoded data by Bit Parity algorithm will be given as an input to AES algorithm for decryption process. It must require secret key to decrypt the data. Data which is generated by AES algorithm will provided SHA-256. It will generate the hash value for SHA-256 which will check the correctness of the data. Download of data is successful in case both hash values i.e. Encrypted hash value and decrypted hash value matches.

IV. IMPLEMENTATION DETAILS AND GRAPH

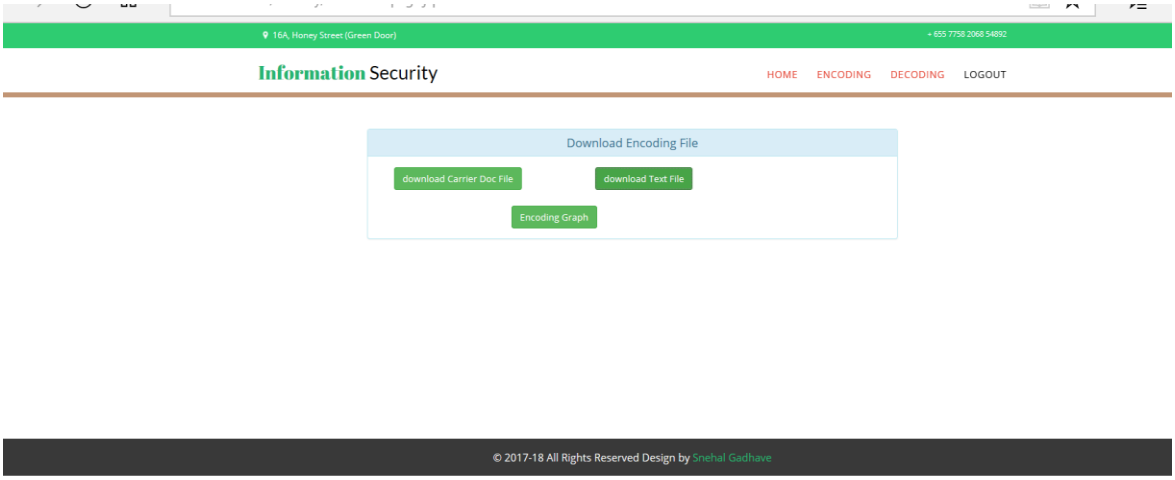
I.Upload document



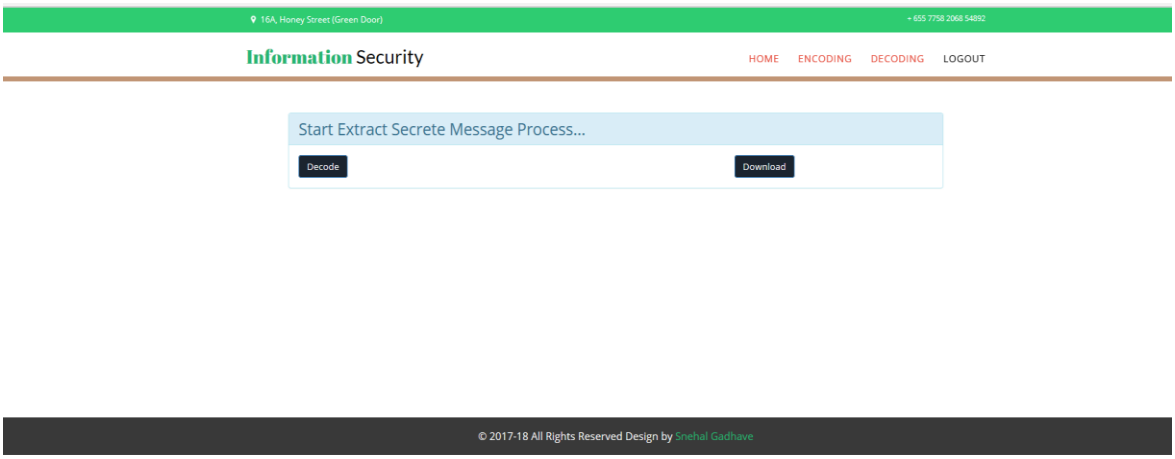
II. Hash value generated:



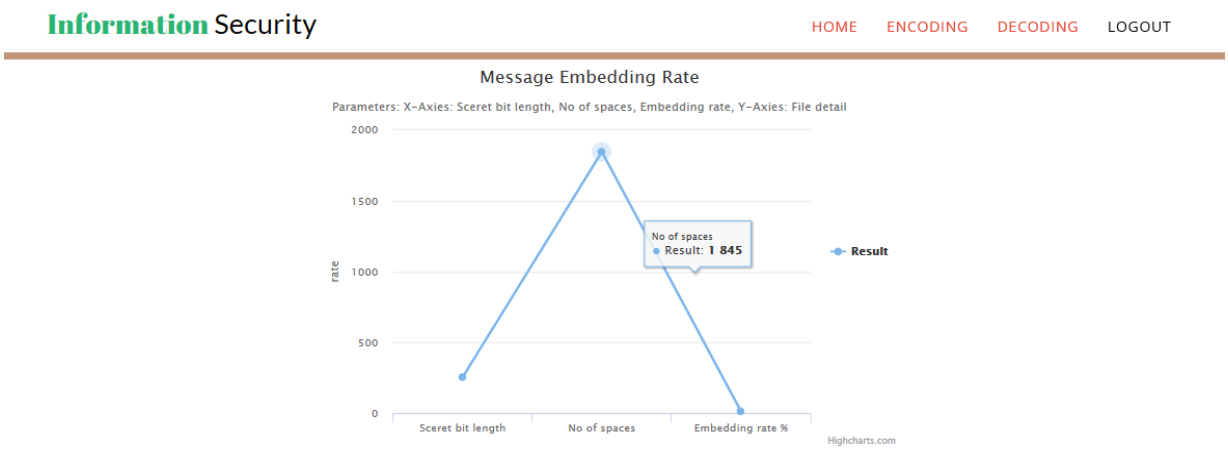
III. Download encoded data



IV. Decode & Download Data



V. Encoding Result



CONCLUSION AND FUTURE WORK

Based on the analysis of the text-based information hiding technology state-of-art, this paper proposed a novel secret information hiding algorithm which is based on the integration of hash function and the invisible ASCII character replacement technology. By introducing a much more complex replace pattern, the generated document can easily resist the brute force secret information detecting methods. By utilizing the collision replacements for the subsequent carrier segments, more available hash values are generated and then higher ER, embedding success and security are all achieved. The experimental results also show that the proposed algorithm is feasible, effective and reliable, robust as well. However, only applicable for the English documents that have plenty of space symbols seriously restricts the applications of the proposed algorithm. So, in the future, we can apply this algorithms to a wider range of languages and all types of file formats as well.

REFERENCES

1. *SHA-3Standard: Permutation-Based Hash and Extendable-Output Functions*, FIPS PUB 202. National Institute of Standards and Technology (NIST). August 2015
2. Pierre 'Moulin, Joseph 'A. *O'Sullivan, Information-theoretic analysis of information hiding*, IEEE Transactions on Information Theory, 49(3), 2003 pp563-593.
3. Niels Provos, Peter Honeyman, *Hide and seek: an introduction to steganography*, IEEE Security Privacy, 1(3), 2003, pp.32ˆa44.
4. 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.
5. Peng Meng, *Research on Linguistic Steganography and Steganalysis*. Ph. D. dissertation, of Science and Technology (China), 2012.
6. Hugo Krawczyk, Mihir Bellare, Ran Canetti, *HMAC: Keyed-Hashing for Message Authentication.* , RFC 2104, February 1997.
7. 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. 96ˆa99.
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. Christian Grothoff, Krista Grothoff, Ryan Stutsman, Ludmila Alkhutova, Mikhail J.Atallah , *Translation based steganography.,* . Journal of Computer Security, 17(3), 2009, pp.269-303.
10. Shufeng Wu. , *A study of information hiding technology*, M.S. dissertation, University of Science and Technology (China), 2003.
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.
12. Ryan Stutsman, Christian Grothoff, Mikhail Atallah, . Lost in just the translation, In: *Proceedings of the 2006 ACM Symposium on Applied Computing (*SAC '06), Dijon, France, April 2006, pp 338-345.
13. AES-The official Advanced Encryption Standard, FIPS PUB 197, *Computer Security Resource Center*. National Institute of Standards and Technology (NIST). Retrieved 26 March 2015.