iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 6 **Issue:** V **Month of publication:** May 2018

DOI: http://doi.org/10.22214/ijraset.2018.5245

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 6 Issue V, May 2018- Available at www.ijraset.com

DNA Based Cryptography and Steganography in Cloud Computing using Socket Programming

Dhanya M S[1], Asha Jose[2], Dincy E B[3]
[1, 2, 3]*Research Scholar, KMP College Of Engineering, Ernakulam, India*

Abstract: ***Cloud computing is the latest technology in the field of distributed computing. Many organizations are unenthusiastic to use cloud services due to data security issues. To address this issue, there have been several approaches applied by various researchers worldwide to strengthen the security of the stored data on cloud computing. In this paper, a blind data hiding hybrid technique is introduced using the concepts of cryptography and Steganography to achieve Triple layer of security in cloud. The proposed method consists of three phases of Encryption: phase one is converting the message to DNA format Followed by applying the Playfair cipher based on DNA and amino acids to encrypt the secret message which generates ambiguity. Phase two is hiding the cipher secret message parts with the ambiguity results from the first phase. Phase three is hiding the corresponding binary of output obtained in phase two using image Steganography. The main Aim is to build a secure and confidential data over cloud.***

Keywords: Cryptography, Steganography, DNA, Playfair, Amino Acid, LSBase, encryption, decryption, cloud computing, LSB method.

I. INTRODUCTION

Cloud computing is the latest technology in the field of distributed computing. It provides various online and on-demand services for data storage, network services, platform services etc. Many organizations are unenthusiastic to use cloud services due to data security issues as the data resides on the cloud service provider's servers. To address this issue, there have been several approaches applied by various researchers worldwide to strengthen security of the stored data on cloud computing. Information security is of increasing importance with the fast developing era as well as its confidentiality. Consequently, high level of security is required as it is a critical feature for thriving networks [1][2][11]. So, the research concerning data hiding techniques has been increased continuously, due to the necessary need for powerful data protection in different applications. Applications such as annotation, ownership protection, copyrighting, military and authentication. Data hiding requires a carrier to hide the data in it such as image, video, and audio as in [1][3][4][5][6].

For achieving maximum protection and powerful security with high capacity and low modification rate, Deoxyribonucleic acid (DNA) is explored as a new carrier for data hiding. A remarkable property of DNA is the capacity in which 106 TB of data can be stored in 1 gm of DNA. However like every data storage device, DNA requires protection through secured algorithm. Various biological properties of DNA sequences can be exploited for obtaining successful secured data embedding process [7][11]. This leads to a new born research field based on DNA computing. The most common and widely used techniques in the communication security and computer security fields are cryptography and steganography [1][3][8][9]. Cryptography includes converting some data to incomprehensible format so that an intended recipient cannot determine its intended meaning [2]. While steganography aims to hide the existence of the message in a different media in order to prevent attracting the attention that the data is there. So, a novel data hiding method is proposed by combining the means of cryptography to encrypt the secret data, steganography to hide the encrypted data and image steganography for hiding the final output to provide Triple layer of security system in cloud computing.

II. RELATED WORKS

In this section, we briefly review some of the recent DNA based steganography techniques for hiding the data. In [10], three data hiding methods were proposed based on DNA and they were considered the main techniques. The first technique is the insertion method by inserting bits from a secret message M randomly in separated positions in a DNA reference sequence. This technique expands the length of the original sequence due to insertion. The second one is the complementary pair method such that the longest complementary pairs in a DNA reference sequence are detected. Then, the message parts are hidden before them so it expands the original sequence's length as well. The last technique is the substitution method that is implemented by substituting some of the DNA nucleotides with other nucleotides based on the secret message bits with no expansion as in [10]. In [5], a data hiding substitution based scheme was proposed by substituting the repeated characters of a DNA sequence. This is done by establishing an

injective mapping between one complementary rule and two secret bits in a message. Complementary rule is the rule that specifies the strand of DNA directly opposite to a specified sequence. This algorithm minimizes the modification rate as a result of substituting the repeated nucleotides of the DNA sequence. This substitution also leads to no expansion in the original reference sequence. However, the modification rate can be very high if the DNA sequence contains a lot of repeated characters. Also, it is not a blind algorithm since the original DNA sequence is required by the receiver to retrieve the secret data [5]. Another idea for data hiding was proposed in [12] through encrypting a secret message using DNA-based Playfair cipher and amino acids.

III.METHODS USED FOR SECURITY OF DATA IN CLOUD COMPUTING

Security methods and technique used by cloud provider are need to regular updation. Many security thread attack for destroy the data on a web or they try to theft the information from the web (cloud). The provider must ensure that their infrastructure is secure and that their clients' data and applications are protected, while the user must take measures to fortify their application and use strong passwords and authentication measures. So we need to develop a secure method. Now days we are using many security algorithms like RSA, ECC, DSA and HASH or some other techniques [17]. Some important and vastly used methods are following:-

A. Cryptography

Cryptography is a method in which we protect data or information and transmit it into an unreadable format. Cryptography play major role to secure ATM transmission, E-commerce, digital media privacy and web data transmission or storage. Modern cryptography work for four major concerns these are non-repudiation, integrity, authentication and confidentiality. In cryptography we use two processes, encryption and decryption.

B. Encryption and Decryption

Encryption is the process to change information (called plain text) into an unreadable secret format (called cipher text). This cipher text could not be easily understood by somebody except authorized parson [17]. Decryption is the process to converting cipher text back into plaintext. The main motive of encryption is to secure the confidentiality of data stored on computer systems or transmitted via one user to other on network (web). In process of data encryption and decryption the user generate a key and use same or different key to decrypt the cipher text or find plain text (information). The goal of every encryption algorithm is to make it as difficult as possible to decrypt the generated cipher text without using the key.

C. Steganography

Steganography is the technique or art of hidden writing. In process of Steganography we are hiding a secret message within an ordinary message (Image, Video and Audio) and the extract message at its destination. We basically study three categories of Steganography, secret key Steganography, public key Steganography and pure Steganography. Its advantage over the cryptography is that the earmarked secret message does not attract attention of third party. So storing and sending data on web is very secure using this technique.

IV. THE PROPOSED TECHNIQUE

The proposed scheme consists of three phases: the first one is converting the secret message to DNA by mapping the binary bits to DNA nucleotides using 2bit and 4bit binary coding rule. Then, the DNA and amino acids based Playfair cipher encrypts the encrypted message. Then in phase2, the ciphered message is hidden in a selected DNA sequence using the LSBase method. Finally, image steganography is performing for hiding the cipher obtained in phase2.So, the first contribution is providing Triple layer security system by developing a hybrid technique to hide the encrypted secret data in DNA results in high security as The second contribution is using 2 bit and 4bit binary coding rule to convert the binary format of a text to DNA that results in increasing the algorithm's cracking probability in obvious way. Finally, the third contribution is the innovation idea 3:1 ratio used in the data hiding technique and also make use of image steganography.

A. Phase I – Data Encryption

In sender's side, the encryption step is preferred before data hiding and image steganography step to avoid hiding the original format of the secret message in the DNA for achieving Triple layer of security: encryption, data hiding and image steganography. The proposed technique uses DNA and amino acids Playfair cipher to encrypt the secret message. The final output obtained is placed inside an image using the technique of image steganography Least Significant Bit method (LSB Method). Conventional Playfair cipher is a symmetric encryption technique that encrypts a text message using a 5*5 table. It is constructed using a secret key word

and the remaining letters of the alphabet that are not included in the key word. Playfair cipher encrypts pairs of letters (digraphs) instead of single letters (monographs) which has advantage in avoiding the attack of the message using frequency analysis of monographs. However, there are severe drawbacks for conventional Playfair cipher that should be taken into consideration [8][13]: It is based on encrypting the message in diagraphs which can be noticed by frequency analysis. Consequently, the attacker may get some information about the data. Another drawback is that it is applied to English alphabet letters only, so, it is unable to encode any special characters or numbers representing equations, numerical data or symbolic data. Also, it is easy to break the system's security now a day's using the advanced computer processors that appear daily. Reference [13] introduced some modification to the conventional Playfair cipher by utilizing some biological concepts as DNA and amino acids to strengthen the ordinary Playfair cipher. In the proposed algorithm, DNA and amino acids based Playfair cipher is applied as the following:

1) Input: Secret Message'M', Secret Key'K'.

2) *Processing:* The Secret Message 'M' is Mapped into its corresponding ASCII and then to Binary using 8-bit coding.

The binary is mapped to DNA nucleotides using 2-bit binary coding rule (BCR) as shown in Table 1.This DNA nucleotide is mapped to amplified message using Table 2.The amplified message obtained will be in a binary format. So this corresponding amplified message is mapped to DNA nucleotides using 4 bit binary coding rule as shown in Table 3.

TABLE 1
2-BITS DNA DIGITAL CODING

Binary Value	DNA digital coding
00	A
01	U or T
10	G
11	C

TABLE 2
KEY COMBINATION

Key Combination	Binary	Key Combination	Binary
AA	0101	GA	1010
AT	0011	GT	0100
AG	0001	GG	1000
AC	0010	GC	1100
TA	0110	CA	1110
TT	1111	CT	1011
TG	0111	CG	0000
TC	1001	CC	1101

The output DNA of the secret message is converted to amino acids according to the new distribution of the alphabet with their corresponding new codons in [13]. The new distribution is derived from the standard universal table of amino acids and their DNA codons representation. Since each amino acid is associated with multiple codons as in [13] and the message is converted from DNA to amino acids. There should be something that refers to the index of each DNA codon corresponding to each amino acid to be able to retrieve the correct codon by the receiver in the decryption phase when amino acids convert to DNA. Playfair cipher is applied using the secret key to encrypt the amino acids form of the secret message formed from the last step into cipher amino acids form. The formed ciphered amino acids are converted back to DNA by selecting the first codon corresponding to each amino acid to form the cipher DNA format of the message. The overall encryption process is illustrated in Fig. 2.

3) Output: Ciphered DNA message.

TABLE 3
4-BITS BINARY CODING RULE

DNA	Binary	DNA	Binary
AA	0000	GG	1000
AC	0001	GA	1001
AG	0010	GC	1010
AT	0011	GT	1011
CC	0100	TT	1100
CA	0101	TA	1101
CG	0110	TC	1110
CT	0111	TG	1111

B. Phase II – Data Hiding

Least significant base is data hiding methodology proposed in [14]. In a DNA sequence each three adjacent nucleotides constitute a unit called codon. LSBase method depends on hiding the secret message bits in the least signification bit of each codon of the reference sequence. Any sequence is a combination of some purine bases (A & G) and pyrimidine bases (C & T). In order to hide the cipher message bits, the following steps are applied:

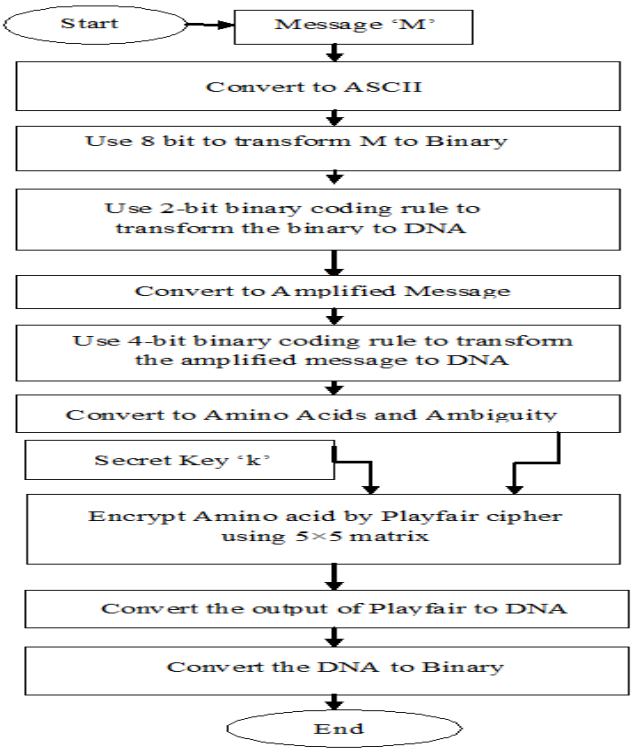


Fig. 2 Data encryption flowchart (Phase I)

1) Input: Ciphered DNA message, ambiguity, and a DNA reference sequence

2) Processing: The formed DNA from the encryption phase is converted into binary again to form cipher binary message by using 4-bits representation binary coding rule as following: The ambiguity as well is converted to binary and since the maximum number of codons corresponding to an amino acid is 4, indexed from 0 to 3 so it can be represented in maximum 2 bits as shown in Fig.4. Select a DNA reference sequence from one of the public databases such as EBI or NCBI and convert it to RNA by substituting each T with U.

Hide cipher binary message and binary value of ambiguity using LSBase method. Since, hiding methodology depends on the message bits and the LSBase of each codon in the DNA. The LSB of DNA reference sequence 'S' is checked and if it is a purine base (A & G), it is substituted by (G) to encode 1 of the secret message or (A) to encode 0. If the LSB of S is a pyrimidine base (C & T), it is substituted by (C) to encode 1 or (U) to encode 0. LSBase algorithm neglects the following codons: UGA, UGG, AUA and AUG during the hiding process since according to the standard distribution of DNA codons to amino acids, Trp and Met amino acids have a single codon which are AUG and UGG respectively [14]. Also, stop has only one codon which is UGA which will be neglected too. Finally it is coded by three codons: AUU, AUC and AUA, so, AUA is neglected and AUU and AUC will be used in data hiding [14]. The complete data hiding scheme is shown in Fig.3.

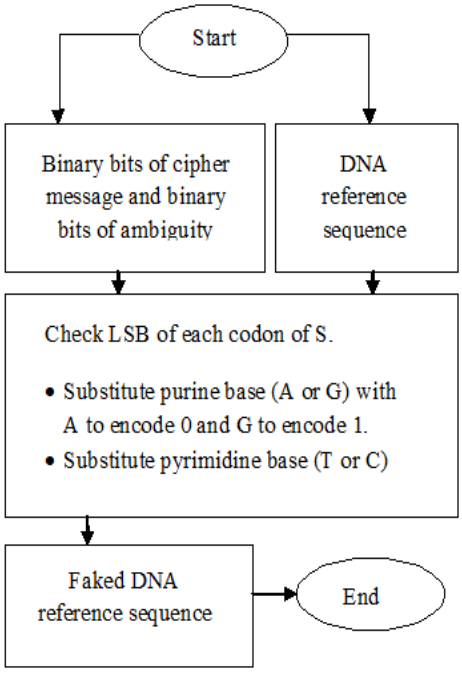


Fig. 3 Data Hiding using LSB method.

3) Output: Output from phase I, not only the cipher binary message but also the ambiguity results from converting DNA format of the message to amino acids. The objective of the proposed method is to hide the secret message and the ambiguity required by the receiver to retrieve the secret message from the DNA without additional information. This because, the ratio of the length of the binary cipher message to the length of the binary ambiguity is 3:1. Hiding the message with the ambiguity in the DNA sequence using 3:1 ratio avoids adding additional data to mark the starting position of the message in the DNA reference sequence and the starting position of the ambiguity as well. Consequently, the data required to be sent is minimized and it is the faked sequence S* only.

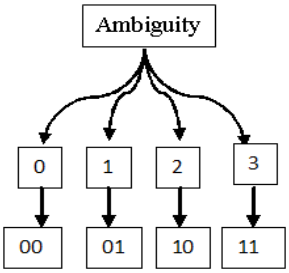


Fig. 4 Conversion rule to convert; Ambiguity to binary.

C. Phase III – Image Steganography

Steganography is one of the most powerful techniques to conceal the existence of hidden secret data inside a cover object. Images are the most popular cover objects for Steganography and in this work image steganography is adopted. The Least Significant Bit (LSB) is one of the main techniques in spatial domain image steganography. The goal of Steganography is to avoid drawing suspicion to the existence of a hidden message. The LSB is the lowest significant bit in the byte value of the image pixel. The LSB based image steganography embeds the secret in the least significant bits of pixel values of the cover image (CVR). The concept of LSB Embedding is simple. It exploits the fact that the level of precision in many image formats is far greater than that perceivable by average human vision. Therefore, an altered image with slight variations in its colours will be indistinguishable from the original by a human being, just by looking at it. In conventional LSB technique, which requires eight bytes of pixels to store 1byte of secret data but in proposed LSB technique, just four bytes of pixels are sufficient to hold one message byte. Rest of the bits in the pixels remains the same as explained in [16]. The following steps are applied:

1) Input: Binary of output obtained in phaseII.

2) Processing: Read the secret and cover image and convert them into gray scale images, then check the size of the secret image with that of the cover image such that size of the secret image should be less than cover image. Encode the secret image into binary using bit gate command and divide it into RGB parts then substitute MSB bits of secret image into LSB bits of cover image.

3) Output: Image containing cipher data.

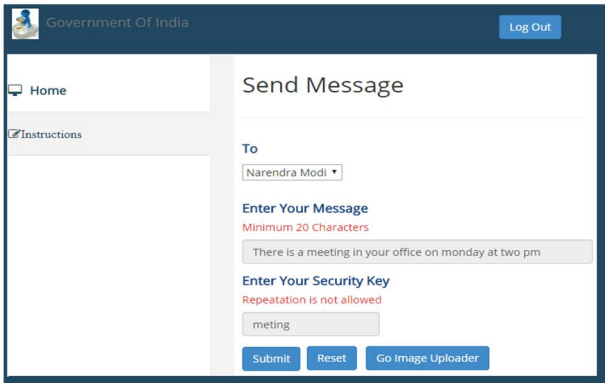


Fig. 5 Secret Message and Key Input Form.

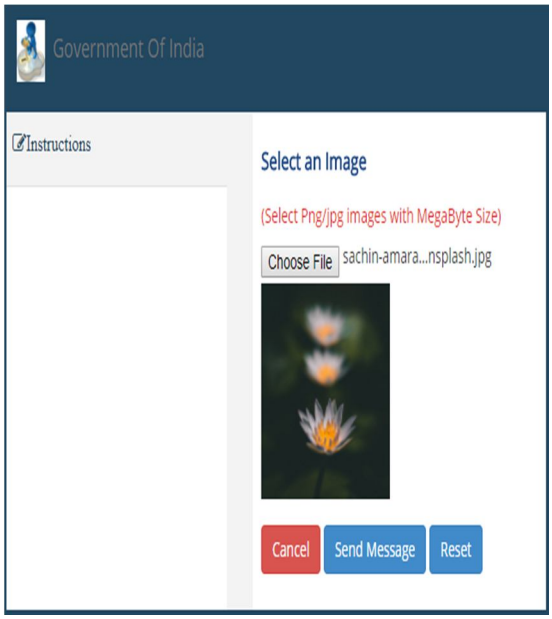


Fig. 6 Image Uploading for performing Image Steganography.

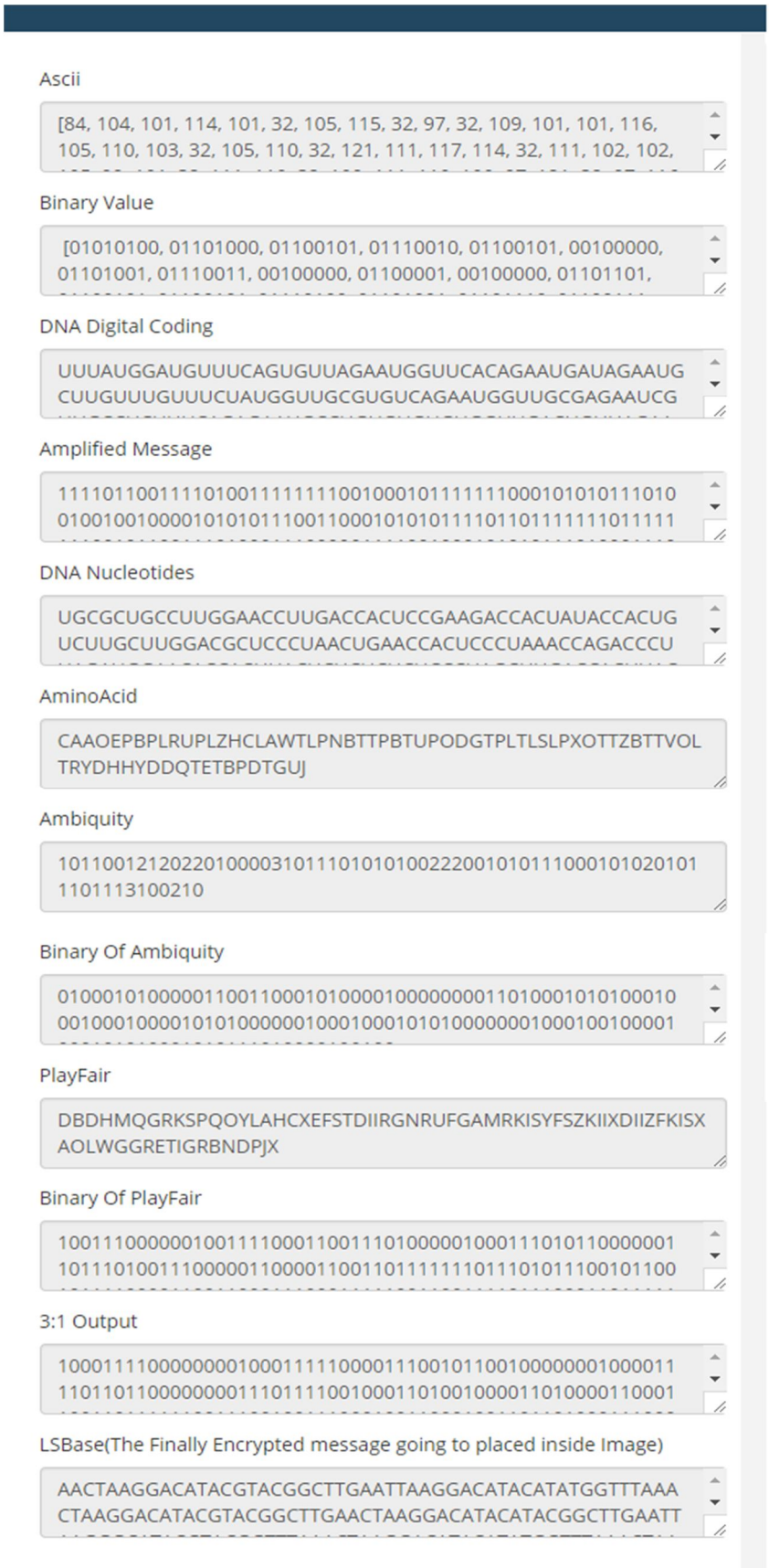


Fig. 7 Step by Step Encryption Process.

D. Data Extraction

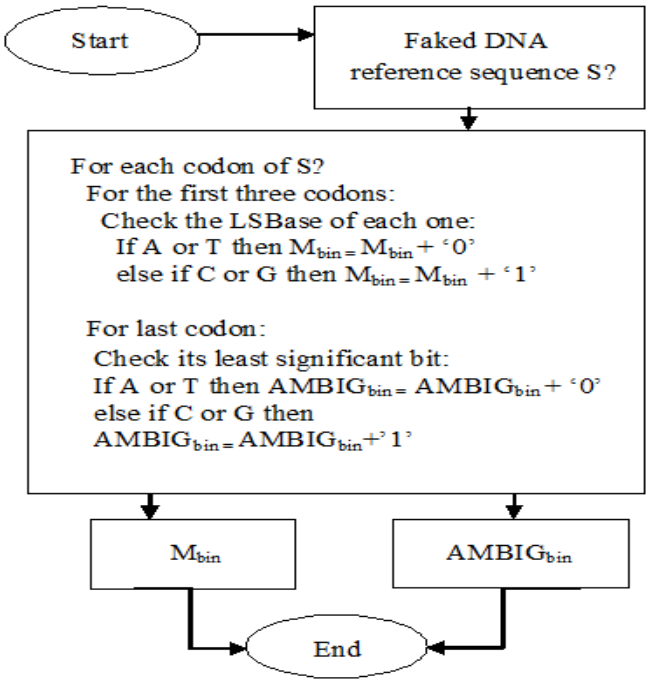
The sender sends the image containing faked DNA sequence to the receiver. Then, the receiver extracts the message which was embedded during embedding process. At first declare a message byte, here the size of the message is 8 bits. Read a pixel from the array starting from address=0.Extract the LSB and replace the i^{th} bit in the message byte where i =1 to 8 Address=address=1. When i =8, a byte is extracted. Repeat for extracting next byte. Finally the faked DNA sequence is extracted from the image.

Then, the receiver applies the Playfair cipher to decrypt the message using the secret key. Both sender and receiver will share the secret key from the beginning. But sharing a secret key may possess a problem since it needs to be interchanged before applying the encryption process. To avoid this problem, the proposed method can be modified to hide the secret key within the faked sequence. The receiver should apply two phases: data extraction and message decryption in order to retrieve the secret data which is contained in the faked DNA sequence.

1) Data Extraction Phase: The extraction process as shown in Fig.8 is simply the inverse of the embedding algorithm where LSBase method is used and S* is divided into codons. Check the least significant base of each codon to retrieve the hidden bits of the secret message. If the LSBase is either 'T'or 'A' then the embedding bit was '0'. If it is 'C'or 'G' then the embedding bit is '1' [14]. Each three extracted consecutive bits by LSBase method are added to the secret message and the next bit is added to the ambiguity of the secret message till binary bits of cipher message (M_{bin}) and binary bits of ambiguity($AMBIG_{bin}$) are completely extracted from S*.

2) Decryption Phase: Decryption is the inverse of the encryption phase as shown in Fig.9, where binary of cipher message (Mbin) is converted to DNA using proposed 4-bits binary coding rule. Then, the ciphered DNA format is converted to amino acids to apply the Playfair cipher on it using the secret key. Decrypted amino acids form is generated from Playfair cipher then binary of ambiguity is converted to decimal digits by mapping each two bit to number. Use each ambiguity number with each amino acid character to retrieve the corresponding codon to this char associated with this ambiguity number. Finally, a sequence of DNA is retrieved, by converting it into amplified message using 4-bits binary coding rule(BCR) then again convert to DNA nucleotides using 2-bit binary coding rule and then to binary and ASCII to get the corresponding plain text which is the original form of the secret data.

Decryption is the reverse operation of encryption. For secret-key encryption, you must know both the key that were used to encrypt the data. For public-key encryption, you must know either the public key (if the data was encrypted using the private key) or the private key (if the data was encrypted using the public key).In this project we are encrypting the secret key using receivers public key and decrypted using receivers private key using RSA algorithm.



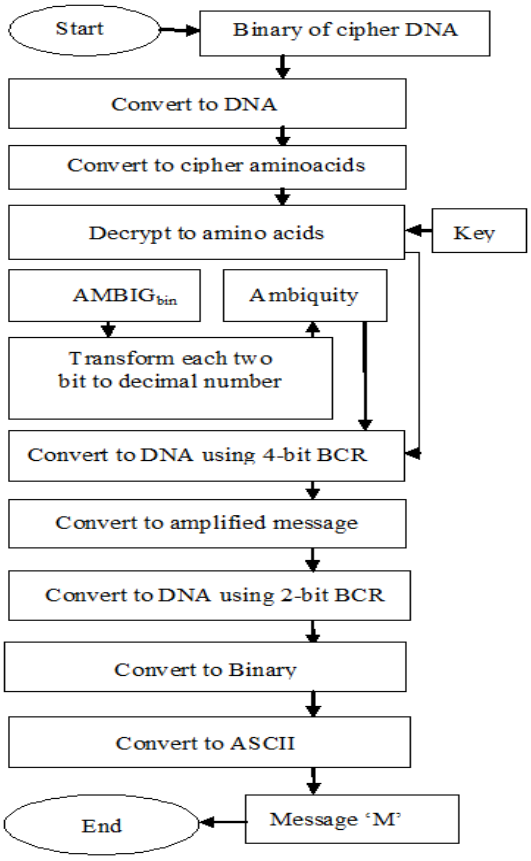


Fig. 9 Decryption Phase at the receiver side.

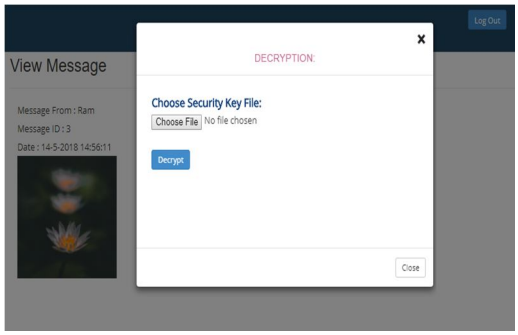


Fig 10. Received Image

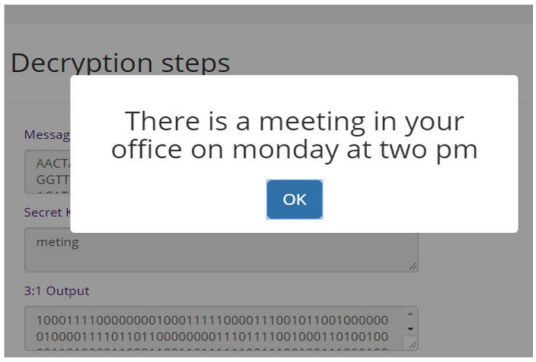


Fig 11. View Original Message

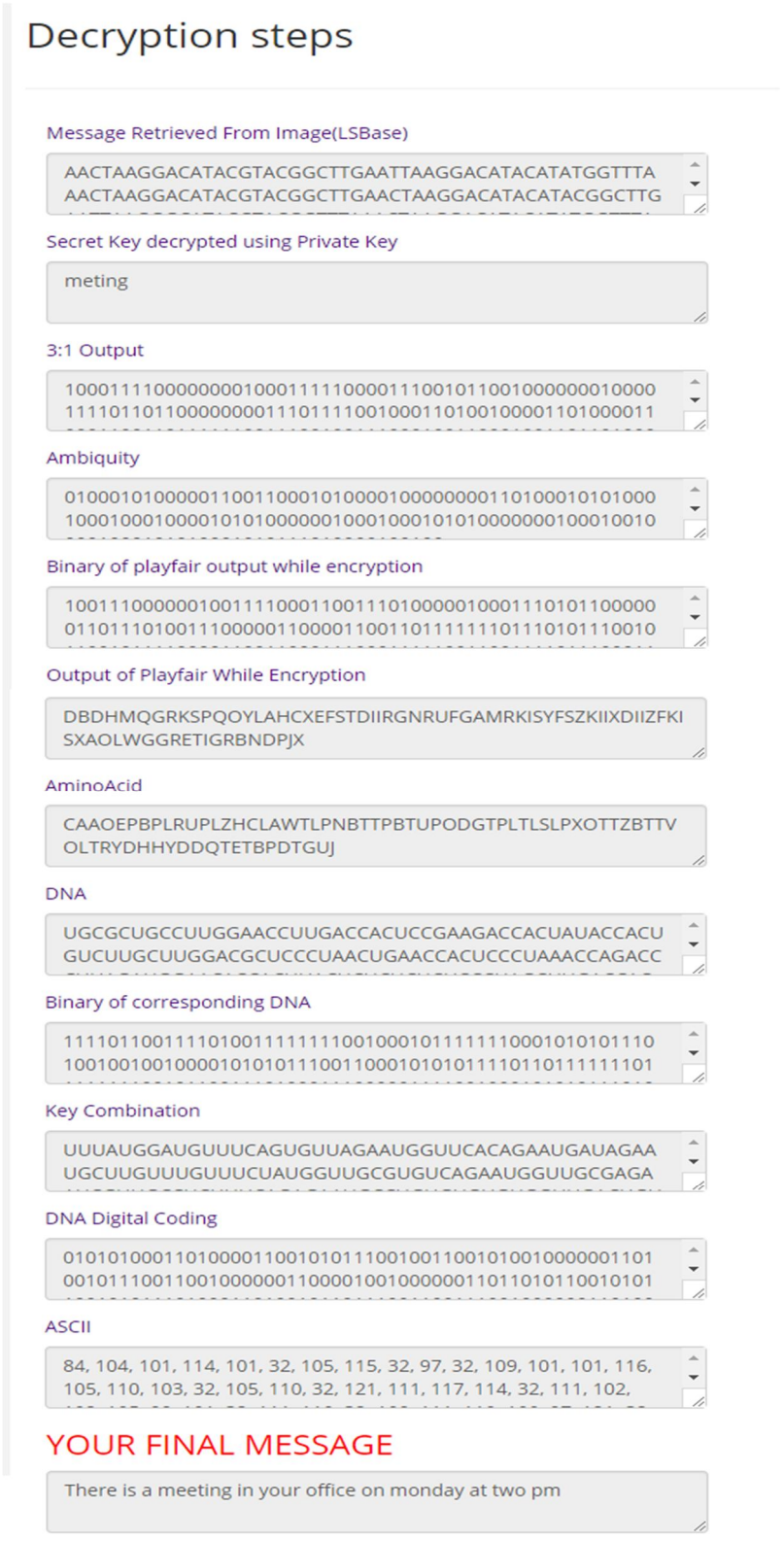


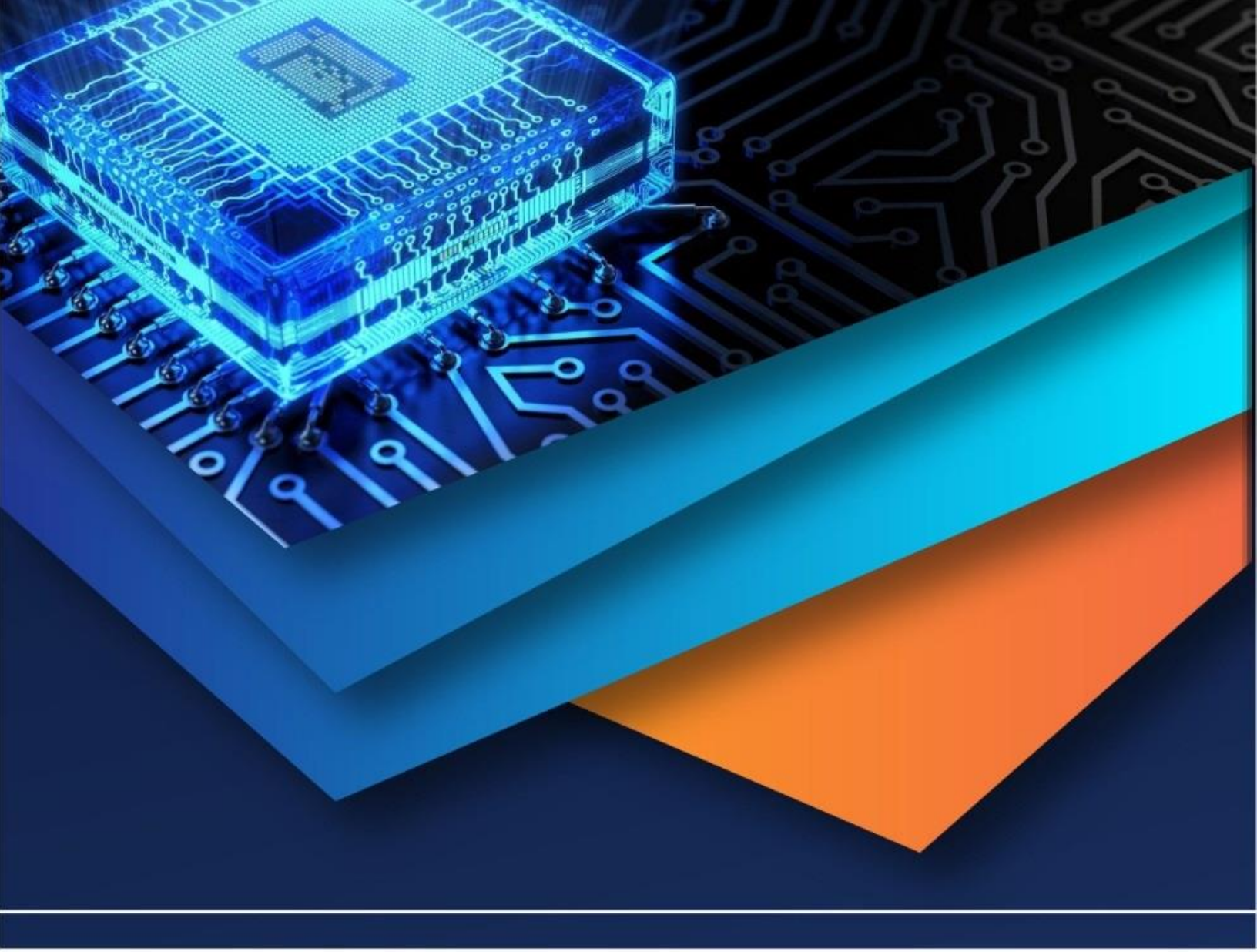
Fig. 12 Step by Step Decryption Process.

V. CONCLUSIONS

Present time of information technology is fully based on online service or web services. Small organization wants to save their money from every time investment on infrastructure development. So they are using online service according to their requirements. In this paper, a data hiding method is proposed by combining the means of cryptography and steganography. Due to using LSB method in hiding the cipher bits of the message and the ambiguity, the proposed algorithm is still blind as the embedded data can be extracted without the need to the original DNA reference sequence. The approach we have used in this paper, will help to make a strong structure for security of data in cloud computing field. As a future work, the proposed method can be modified to increase the hiding capacity of the DNA sequence by using some other techniques with cryptography and make use of fully homomorphic encryption for encrypting the image that helps to secure the information on cloud.

REFERENCES

[1] G. Hamed et al, "DNA Based Steganography: Survey and Analysis for Parameters Optimization," in Applications of Intelligent Optimization in Biology and Medicine, Springer, 2015, ISSN: 1868–4394, pp. 47–89.
[2] B.A.Mitras and A.K. Abo, "Proposed Steganography Approach using DNA Properties," International Journal of Information Technology and Business Management, ISSN: 2304–0777, vol. 14, Issue No. 1, pp. 96– 102, June 2013.
[3] M. Skariya and M. Varghese, "Enhanced Double Layer Security using RSA over DNA based Data Encryption System," International Journal of Computer Science & Engineering Technology (IJCSET), ISSN: 2229– 3345, vol. 4, Issue No. 06, pp. 746–750, Jun 2013.
[4] Y. A. Yunus, S. Ab Rahman and J. Ibrahim, "Steganography: A Review of Information Security Research and Development in Muslim World," American Journal of Engineering Research (AJER), ISSN: 2320–0936, vol. 02, Issue No. 11, pp. 122–128, 2013.
[5] C. Guo, C. Change and Z. Wang, "A New Data Hiding Scheme based on DNA Sequence," International Journal of Innovative Computing, Information and Control, ISSN:1349–4198, vol.8, IssueNo. 1, pp.139– 149, Jan 2014.
[6] I.K..Maitra, "Digital Steganalysis: Review on Recent Approaches," Journal of Global Research in Computer Science, ISSN: 2229–371X, vol. 2, Issue No. 1, pp. 1–5, Jan 2011.
[7] J. Taur, H. Lin, H. Lee and C. Tao, "Data Hiding in DNA Sequences based on Table Lookup Substitution," International Journal of InnovativeComputing, Information and Control, ISSN: 1349–4198, vol. 8, Issue No. 10, pp. 6585–6598, Oct. 2012.
[8] Atito, A. Khalifa and S. Z. Rida, "DNA-based Data Encryption and Hiding using Playfair and Insertion Techniques," Journal of Communications and Computer Engineering, ISSN: 2090–6234, vol. 2, Issue No. 3, pp. 44–49, 2012.
[9] K. Kaundal and A. K. Verma, "DNA based Cryptography: A Review," International Journal of Information and Computation Technology, ISSN: 0974–2239, vol. 04, Issue No. 7, pp. 693–698, 2014.
[10] H.J. Shiu, K.L. Ng, J.F. Fang, R.C.T. lee and C.H. Huang, "Data Hiding Methods based upon DNA Sequences," Journal of Information Sciences: an International Journal, vol. 180, Issue No. 11, pp. 2196– 2208, June 2010.
[11] M. R. N. Torkaman, N. S. Kazazi and A. Rouddini, "Innovative Approach to Improve Hybrid Cryptography by using DNA Steganography," International Journal of New Computer Architectures and their Applications (IJNCAA), ISSN: 2220–9085, vol. 2, Issue No. 1, pp. 224– 235, 2012.
[12] Khalifa and A. Atito, "High-Capacity DNA-based Steganography," in Informatics and Systems (INFOS), 8th International Conference on 2012, May 2012, pp. BIO–76.
[13] M. Sabry, M. Hashem, T.Nazmy and M. E. Khalifa, "A DNA and Amino Acids-based Implementation of Playfair Cipher," International Journal of Computer Science and Information Security, ISSN: 1947–5500, vol.8, Issue No. 3, pp. 129–136, 2010.
[14] Khalifa, "LSBase: A key Encapsulation Scheme to Improve Hybrid Crypto-systems using DNA Steganography," in 8th International Conference on Computer Engineering & Systems (ICCES), Cairo, Egypt, Nov. 2013, pp. 105–110.
[15] NCBI Database, Bank for real DNA reference sequences, http://www. ncbi.nlm.nih.gov/
[16] Champakamala .B.S, Padmini.K, Radhika .D. K," Least Significant Bit algorithm for image steganography, " International Journal of Advanced Computer Technology (IJACT), ISSN: 2319-7900, Vol. 3, IssueNo. 4.
[17] Marwaha, R.Bedi, "Applying Encryption Algorithm for Data Security and Privacy in Cloud Computing, ", IJCSI International Journal of Computer Science Issues, Vol. 10, IssueNo. 1, P. 367-370, January 2013.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY