A Major Project Report On

IMAGE ENCRYPTION USING DNA AND BLOCK BASED CIPHER

Submitted in partial fulfilment of the requirements

For the award of the degree of

MASTER OF TECHNOLOGY

in

COMPUTER SCIENCE & ENGINEERING

By

Praveen Kumar Singhal

(Roll No. 2K13/CSE/16)

Under the guidance of

Mr. Manoj Kumar

Associate Professor

Department of Computer Science & Engineering

Delhi Technological University, Delhi

Department of Computer Science & Engineering

Delhi Technological University, Delhi

2013-2015

DELHI TECHNOLOGICAL UNIVERSITY

CERTIFICATE

This is to certify that the project report entitled "**Image Encryption using DNA and Block Based Cipher"** is a bona fide record of work carried out by Praveen Kumar Singhal (2K13/CSE/16) under my guidance and supervision, during the academic session 2013-2015 in partial fulfilment of the requirement for the degree of Master of Technology in Software Engineering from Delhi Technological University, Delhi.

To the best of my knowledge, the matter embodied in the thesis has not been submitted to any other University/Institute for the award of any Degree or Diploma.

Manoj Kumar
Associate Professor
Department of Computer Science & Engineering
Delhi Technological University
Delhi

DELHI TECHNOLOGICAL UNIVERSITY

ACKNOWLEDGEMENT

With due regards, I hereby take this opportunity to acknowledge a lot of people who have supported me with their words and deeds in completion of my research work as part of this course of Master of Technology in Software Engineering.

To start with I would like to thank the almighty for being with me in each and every step of my life. Next, I thank my parents and family for their encouragement and persistent support.

I would like to express my deepest sense of gratitude and indebtedness to my guide and motivator, **Mr. Manoj Kumar**, Associate Professor, Department of Computer Science and Engineering, Delhi Technological University for his valuable guidance and support in all the phases from conceptualization to final completion of the project.

I wish to convey my sincere gratitude to our Head of Department, and all the faculties and Ph.D. Scholars of Computer Engineering Department, Delhi Technological University who have enlightened me during my project.

Praveen Kumar Singhal
2K13/CSE/16

TABLE OF CONTENTS

CHAPTER 1

List of Figures

List of Tables

ABSTRACT

The blowfish algorithm is a block based encryption symmetric key encryption algorithm and has many advantage in symmetric data encryption. But when we encrypt an image which has large block of single colour than after encryption it convert this block to same but different colour so it may be easily to understand that this large block has single colour so any has crack the image. So we add DNA cryptography to improve the encryption algorithm. DNA cryptography encrypt data at bit level. In this thesis logistic map is used to generate the DNA sequence. To improve the randomness in generating the DNA sequence, we generate the initial value of logistic map using blowfish key, so every time DNA sequence is different on different blowfish key. Initial value of logistic map is calculated using hash function. In the first stage of this approach the image will be converted into DNA sequence and another DNA sequence will be generated using logistic map then DNA addition operation is performed. After anti clockwise rotation is performed on output then output is feed to blowfish algorithm. The output is measured for the security level based on correlation, histogram and entropy. The experiment results have shown that the combination technique resulted in a higher entropy value and lower correlation and more uniform histogram.

Chapter-1

Introduction

In this chapter, we have discussed cryptography and security and security principals. There are two types of cryptography: secret key cryptography and public key cryptography. In secret key cryptography same key is used for both encryption and decryption. In public key cryptography, each user has a public and private key.

1.1 What is Cryptography

Cryptography is the discipline of cryptography and cryptanalysis and their interaction. The word “cryptography” is derived from the Greek words “Kryptos” meaning concealed, and “graphien” meaning to inscribe. It is the science of keeping secrets secret. One objective of cryptography is protecting a secret from adversaries. Professional cryptography protects not only the plain text but also the key and more generally tries to protect the whole cryptosystem. Cryptographic primitives can be classified into two classes: keyed primitives and non-keyed primitives as shown in figure [1]. Providing confidentiality using encryption algorithm is the fundamental task of cryptography. Encryption (also called as enciphering) is the process of scrambling the contents of a message or file to make it unintelligible to anyone not in possession of key “key” required to unscramble the file or message. Providing confidentiality is not the only objective of cryptography. Cryptography is also used to provide solutions for other problems: Data integrity, Authentication, Non-repudiation [2].

Encryption methods can be divided into two categories: substitution ciphers and transposition ciphers. In a substitution cipher, the letters of the plaintext are replaced by other letters or by symbols or numbers. Replacing plaintext bit pattern with ciphertext bit patterns is involved in substitution when plain text is viewed as a sequence of bits. Substitution ciphers preserve the order of plaintext symbols but disguise them. Transposition ciphers, do not disguise the letters, instead reorder them. This is achieved by performing some sort of permutation on the plaintext letters.

There are two types of encryption: symmetric (private/secret) encryption key and asymmetric (public) key encryption.

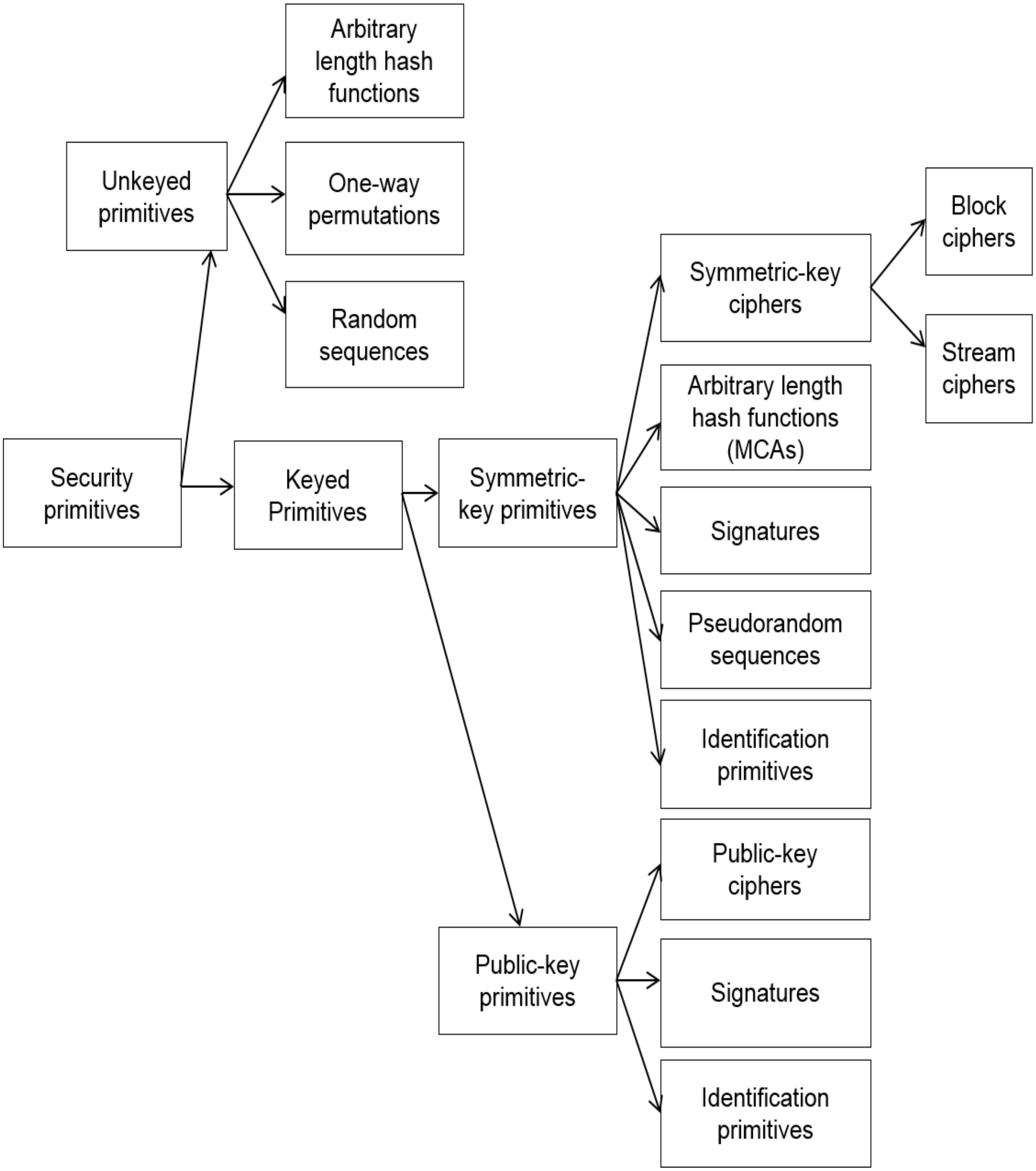


Fig.1: Cryptographic Primitives

1.2 Conventional Encryption Model

A conventional encryption model can be illustrated as assigning X_p to represent the plaintext message to be transmitted by the originator. The parties involved select an encryption algorithm represented by E. The parties agree upon the secret key represented by K. The secret key is distributed in a secure manner represented by SC. Conventional encryption's effectiveness rests on keeping the secret. Keeping the key secret rests in a large on key distribution methods. When E process X_p and K, X_c is derived. X_c represents the ciphertext output, which will be decrypted by the recipient. Upon receipt of X_c, the recipient uses a decryption algorithm represented by D to process X_c and K back to X_p. This is represented in the figure. With conventional encryption, the secrecy of the encryption and decryption algorithm is not needed.

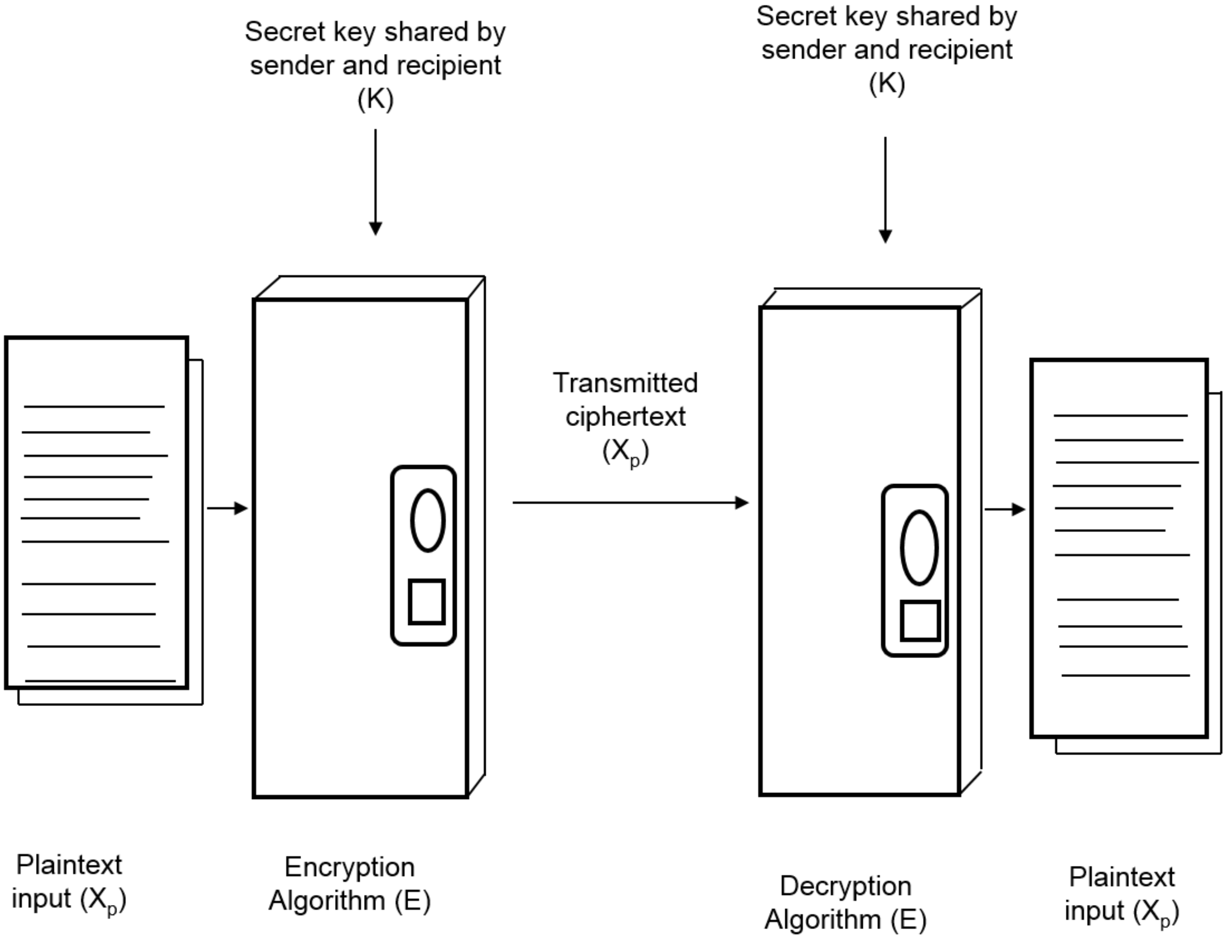

Secret key shared by sender and recipient (K)
Secret key shared by sender and recipient (K)
Transmitted ciphertext (X_p)
Plaintext input (X_p)
Encryption Algorithm (E)
Decryption Algorithm (E)
Plaintext input (X_p)

Fig.2: Simplified Model of Conventional Encryption

1.3 Cryptanalysis

Code making involves the creation of encryption products that provide protection of confidentiality. Defeating this protection by some men's other than the standard decryption process used by an intended recipient is involved in code breaking. Five scenarios for which code breaking is used. They are selling cracking product and services, spying on opponents, ensure accessibility, pursuing the intellectual aspects of code breaking and testing whether one's codes are strong enough. Cryptanalysis is the process of attempting to identify either the plaintext X_p or the key K. Discovery of the encryption is the most desired one as with its discovery all the subsequent messages can be deciphered. Therefore, the length of the encryption key, and the volume of the computational work necessary provides for its length i.e. resistance to breakage. The protection gets stronger when key size increases but this requires more brute force. Neither encryption scheme conventional encryption nor the public key encryption is more resistant to cryptanalysis than the other.

1.4 Cryptographic Goals

However, there are other natural cryptographic problems to be solved and they can be equal if not important depending on who is attacking you and what you are trying to secure against attackers. Privacy, authentication, integrity and non-repudiation are the cryptographic goals covered in this text.

- **Confidentiality:** Preserving authorized restrictions on information access and disclosure, together with means for shielding personal secrecy and copyrighted material. A damage of privacy is the illegal disclosure of information.
- **Integrity:** Integrity ensures that information which we send, during network it cannot be modified by the intruder and also ensure the information authenticity and non-repudiation. When some other than authorized person modify the information then we losses integrity.
- **Availability:** Availability means at any time and at any location, access and use of information. If information is not available at any time or at any location then we losses availability.

To represent the complete picture, some additional concepts are needed. Two of the commonly are:

- **Authenticity:** Authenticity means only sender and receiver of a message can identify the message. In this, both parties have to trust each other for ensuring authenticity.
- **Accountability:** The security goal that generates the requirement for actions of an entity to be traced uniquely to that entity.

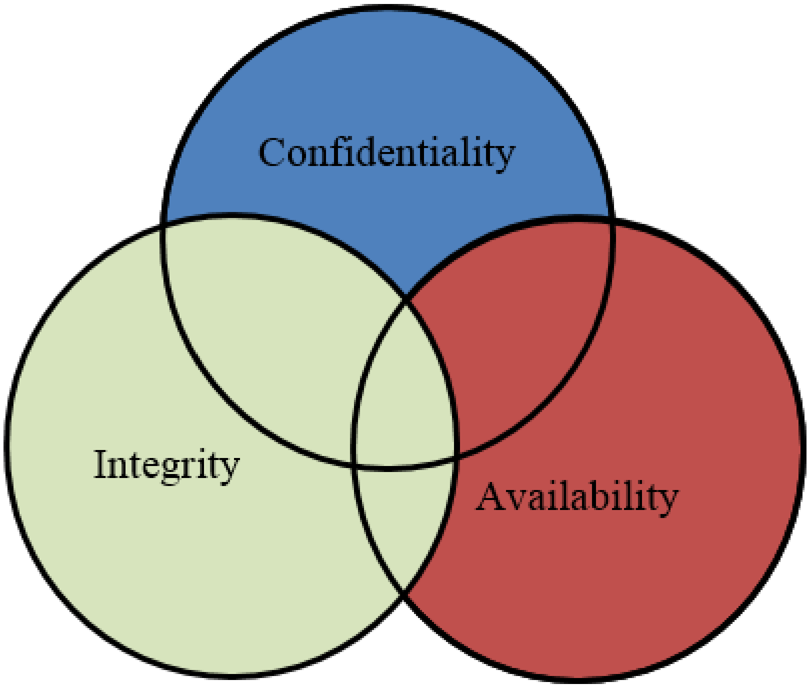


Fig.3: Cryptography Goals

Generally there are two types key present

- Symmetric-key
- Asymmetric-key

1.5 Symmetric Key Encryption

Symmetric encryption is a universal technique for transmitting data and providing confidentiality of data. Symmetric encryption is also known as single-key encryption Symmetric encryption is also known as single-key means it uses the same key for encryption and decryption of the information. Countless individuals and groups, from Julius Caesar to the German U-boat force to present-day diplomatic, military and commercial users, use symmetric encryption for secret communication. It remains by far the more widely used of the types of encryption. A symmetric encryption scheme has five ingredients as follows-

- **Plaintext:** This is the original data or message is taken as input for the algorithm.
- **Encryption algorithm:** The encryption algorithm encrypts data by performing various transformations and substitutions on the plaintext input.
- **Secret key:** It is input to the encryption algorithm. The exact transformations and substitutions performed by the encryption algorithm mainly depend on the input key.
- **Ciphertext:** This is the output produced by the encryption algorithm that is scrambled the message. It depends on the plaintext input and the secret key. For a given input message, two different keys will produce two different ciphertexts.
- **Decryption algorithm:** This is reserve process of the encryption algorithm. It takes the ciphertext and secret key and produces the original plaintext.

Symmetric key encryption is shown in fig.

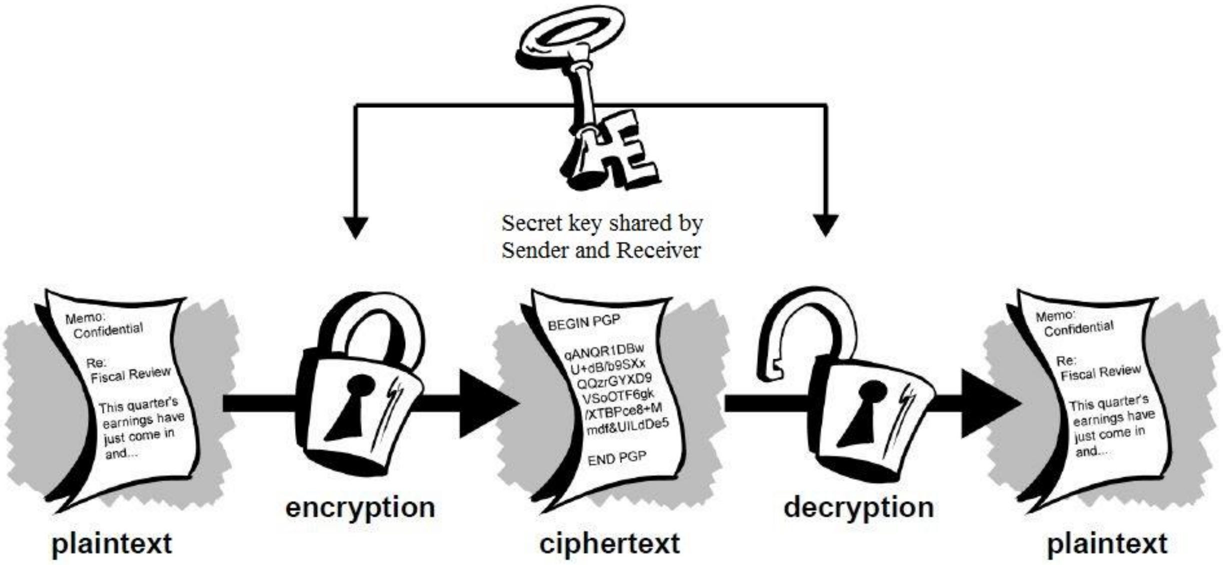


Fig.4: Symmetric key encryption

There are two necessities for protected use of symmetric encryption:

- We need a good encryption algorithm.
- Sender and receiver must have secured obtained, & keep secure, the secret key.

1.5.1 Stream Ciphers

The stream ciphers encrypt data by generating a key stream from the key and performing the encryption operation on the key stream with the plaintext data. The key stream can be any size that matches the size of the plaintext stream to be encrypted. The i^{th} key stream digit only depends on the secret key and on the (i-1) previous plaintext digits. Then, the i^{th} ciphertext digit is obtained by combining the i^{th} plaintext digit with the i^{th} key stream digit [2]. One desirable property of a stream cipher is that the ciphertext be of the same length as the plaintext. Thus, a ciphertext output of 8 bits should be produced by encrypting each character if 8-bit characters are being transmitted. Transmission capacity is wasted if more than 8 bits are produced. However, stream ciphers are vulnerable to attack if the same key is used twice or more.

1.5.2 Block Ciphers

A block ciphers fragments the message into blocks of a predetermined size and performs the encryption function on each block with the key stream generated by cipher algorithm. The size of each block should be fixed, and leftover message fragments are padded to the appropriate block size. Block ciphers differ from stream ciphers in that they encrypted and decrypted information in fixed size blocks rather than encrypting and decrypting each letter or word individually. A block ciphers passes a block of data or plaintext through its algorithm to generate a block of ciphertext.

1.6 Asymmetric Key Cryptosystems

In Asymmetric Key Cryptosystems, two different keys are used: a secret key and a public key. The secret key is kept undisclosed by the proprietor and public key is openly known. The system is called “asymmetric” since the different keys are used for encryption and decryption, the public key and private key.

If data is encrypted with a public key, it can be decrypted only by using the corresponding private key. Public Key Encryption shown in the fig.

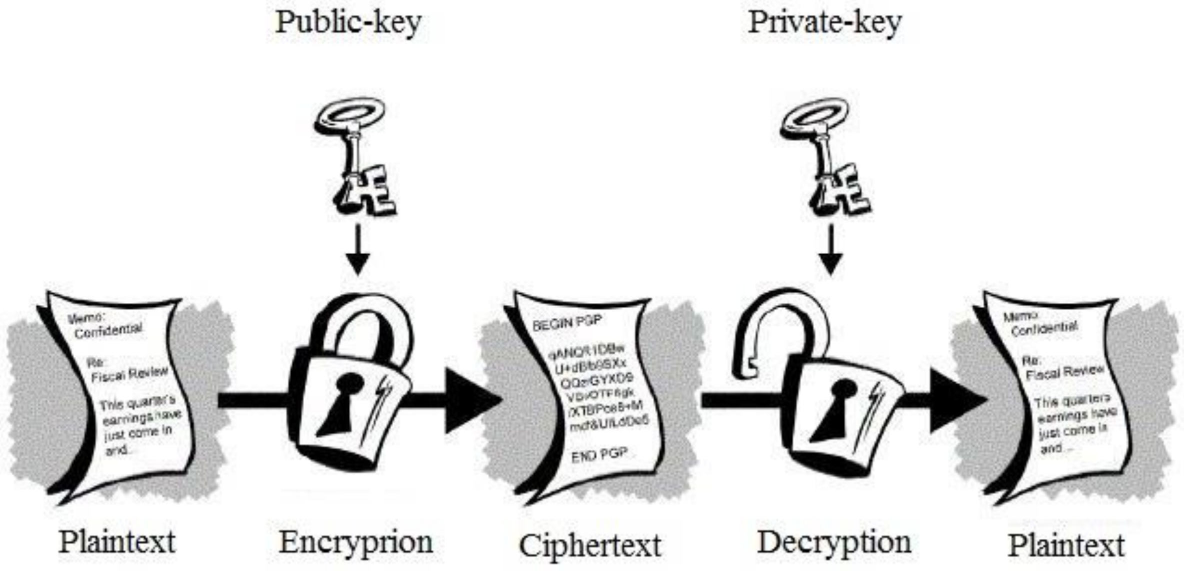


Fig.5: Public Key Encryption

1.7 Classical Encryption Techniques

The technique enables us to illustrate the basic approaches to conventional encryption today. The two basic components of classical ciphers are transposition and substitution. A Combination of both substitution and transposition is described in others systems.

1.7.1 Substitution Techniques

In this technique letters of plaintext message are placed by symbols and numbers. If the plaintext is in the form of sequences of bits, then substituting plaintext bit patterns with ciphertext bit patterns.

1.7.2 Transposition Techniques

Transposition instantly moves the position around within it but does not alter any of the bits in the plaintext. If the resultant ciphertext is then put through more transpositions, the end result has increasing security.

1.8 Research Motivations

Most of the image encryption algorithms were designed in the mid-1990s. According to the Bourbakis and Maniccam (2004), there are two categories of image encryption algorithm: (a) Non-chaos selective methods and (b) chaos based selective or non-selective methods (Manicca and Bourbakis, 2004). However, most of encryption algorithms are used for a specific image format, either compressed format or uncompressed format. There are procedures that offer light encryption as well as a strong form of encryption methods. But some of the image algorithms have different modes ranging from strong to degradation encryption. According to Borke (2005), the user has an option to choose which method is best suitable for image security based on some properties of the image.

Image encryption has applications in multimedia systems, medical and military imaging systems, internet communication. All multimedia data have some characteristics like high correction among pixels and high redundancy. Thus, to protect our data, various techniques are used so that image data is not used by the unauthorized person.

The motivation behind this search is the ever-increasing need for hard-to –break encryption and decryption algorithm as the network technologies evolve. Proposing this algorithm we believe that it will reduce the relationship among image pixel by increasing the entropy value and decreasing the correlation coefficient value.

1.9 Goal, Scope and Objectives of the Research

There are so many algorithms present to encrypt and decrypt the data for security purpose in cryptography. Blowfish encryption method is one of the block based encryption algorithms. Bruce Schneier in 1993 first developed the blowfish ciphers algorithm. In this algorithm, DNA cryptography is used with Blowfish cipher algorithm to improve the security level of cipher images. In novel proposed cipher algorithm is quite robust as for as cryptanalysis is concerned. Proposed cipher algorithm is comparatively better approach as compared to blowfish cipher in terms of correlation and entropy of encrypted images. The objective of this algorithm is to overcome the drawbacks of blowfish algorithm and improve the security.

1.10 Problem definition

In this thesis, to enhance the result of image encryption we use DNA cryptography with blowfish encryption cipher. Only using blowfish cipher we cannot get good encrypted image. Hare, we also use image rotation to scrambling the original image. Hare we use image rotation at the bit level (we rotate at the binary level of the image) and at pixel level to improve encryption process.

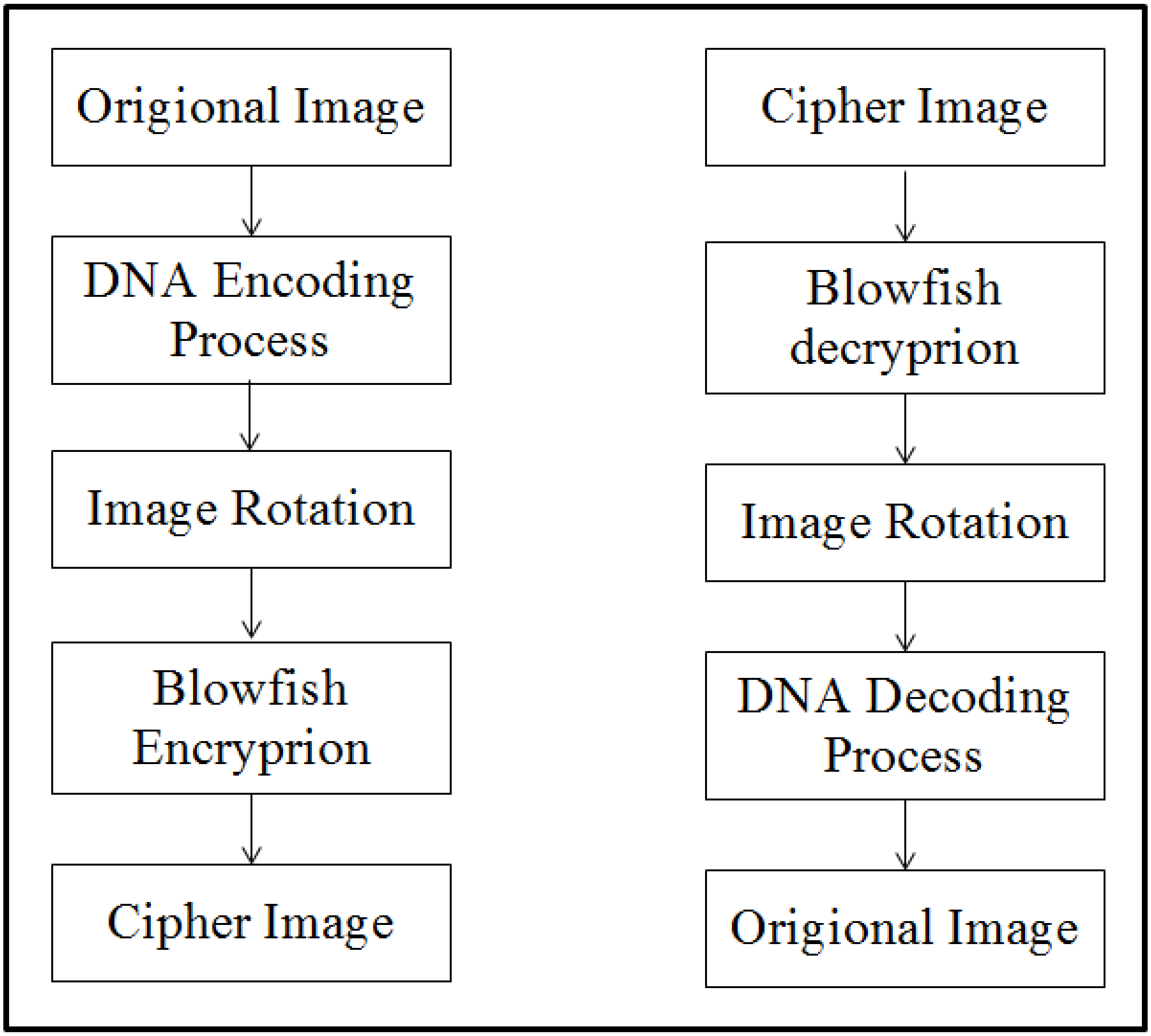


Fig.6: General block diagram of the proposed technique

Experiment result are given to demonstrate the proposed algorithm that are significantly more effective in the encryption quality of the image than using only blowfish cipher and it is also compared with some image encryption algorithm. A general block diagram of proposed technique is shown in the figure.

1.11 Dissertation Structure

The disposition of the dissertation explains the documentation of the dissertation work chapter by chapter.

Chapter-1 Introduction: This chapter is the introduction part of the dissertation work. It contains the description of the cryptographic background of the carried out research, an introduction about the cryptographic methods and security principle, the research problem, the research objective, the type of methodology used in the dissertation work, structure of the dissertation report.

Chapter-2 Research Background: This chapter contains theory study of the Blowfish cipher, MD5 hash function, DNA cryptography and chaotic map.

Chapter-3 Literature Survey: The second chapter of this report consists of the literature study of DNA cryptography.

Chapter-4 Proposed Method: The third chapter of the dissertation report contains the description of the implementation part of the algorithm.

Chapter-5 Experiment and Result analysis: The algorithm is implemented for different images and compares the result with existing algorithm in terms of correlation, entropy, MSE and NPCR in this chapter.

Chapter-6 Conclusion: In this chapter the concluding explanations of the algorithm are given.

Chapter 2

Research Background

2.1 Blowfish

Blowfish cipher is a symmetric key block cipher that is effectively used for encryption of data. Blowfish cipher was designed by Bruce Schneier in 1993. This algorithm is a fast, free alternative to existing cipher algorithm to encrypt data. Blowfish cipher is a license-free and free available to every user. Blowfish key size is not fixed, it varies from 32 bits to 448 bits, so blowfish is ideal for securing data [5].

Blowfish Algorithm uses the Feistel Network, iterate its function (F) 16 times for encryption of data. Block size in blowfish cipher is 8 bytes, and the key length varies 32 bit to 448 bits. Blowfish have complex initialization phase before encrypting the data.

2.1.1 Feistel Networks

A Feistel network usually called as function F is used in block cipher to generate the permutation of a given block data. The name of this network is Feistel after it is designed by Horst Feistel.

The general workflow of the network as given below:

- First it divides every data into, two parts one is left and another is right half, each has equal size.
- After the process the both left and right half, right half become simply a new left half.
- The right half and key are taken in function and result is XORed with left half and the final result stored in the right half.

2.1.2 Blowfish Algorithm

In blowfish cipher, every input data is 64 bit (8-byte) long and have variable length key.

- Algorithm has 64-bit input data as d.
- Divide d into left (dL) and right (dR) half, each of 32 bits.

- For I = 1 to 16:
    - dL = dL XORed Pi
    - dR = F(dL) XORed dR
    - Swap dL and dR
- After last iteration, we undo the last swap.
- dR = dR XORed P17
- dL = dL XORed P18
- Combine dL and dR again.
- Function F:
    - Divide dL into equal parts each of size 8 bits as a, b, c, and d.
    - F(dL) = ((S1[a] + S2[b] mod 2^{32}) XOR S3[c]) + S4[d] mod 2^{32}

Generally blowfish algorithm divides into two parts. First is key expansion part, in this part variable length input key is used to generate the several sub-keys and second is data encryption part, in which algorithm take input data and key and encrypt this data. This algorithm does not have any complex operation, it have only simple addition and XORed operation.

2.1.3 Sub-keys

Sub-keys are generated in key expansion part before any encryption is performed. Sub-keys has 32 P-array each of size 32 bits and four S-box each contains 256 entries each of which has size 32 bits.

The P-array:

P1, P2… P18.

S-box:

S1 [0], S1 [1] ……. S1 [255],
S2 [0], S2 [1] ……. S2 [255],
S3 [0], S3 [1] ……. S3 [255],
S4 [0], S4 [1] ……. S4 [255].

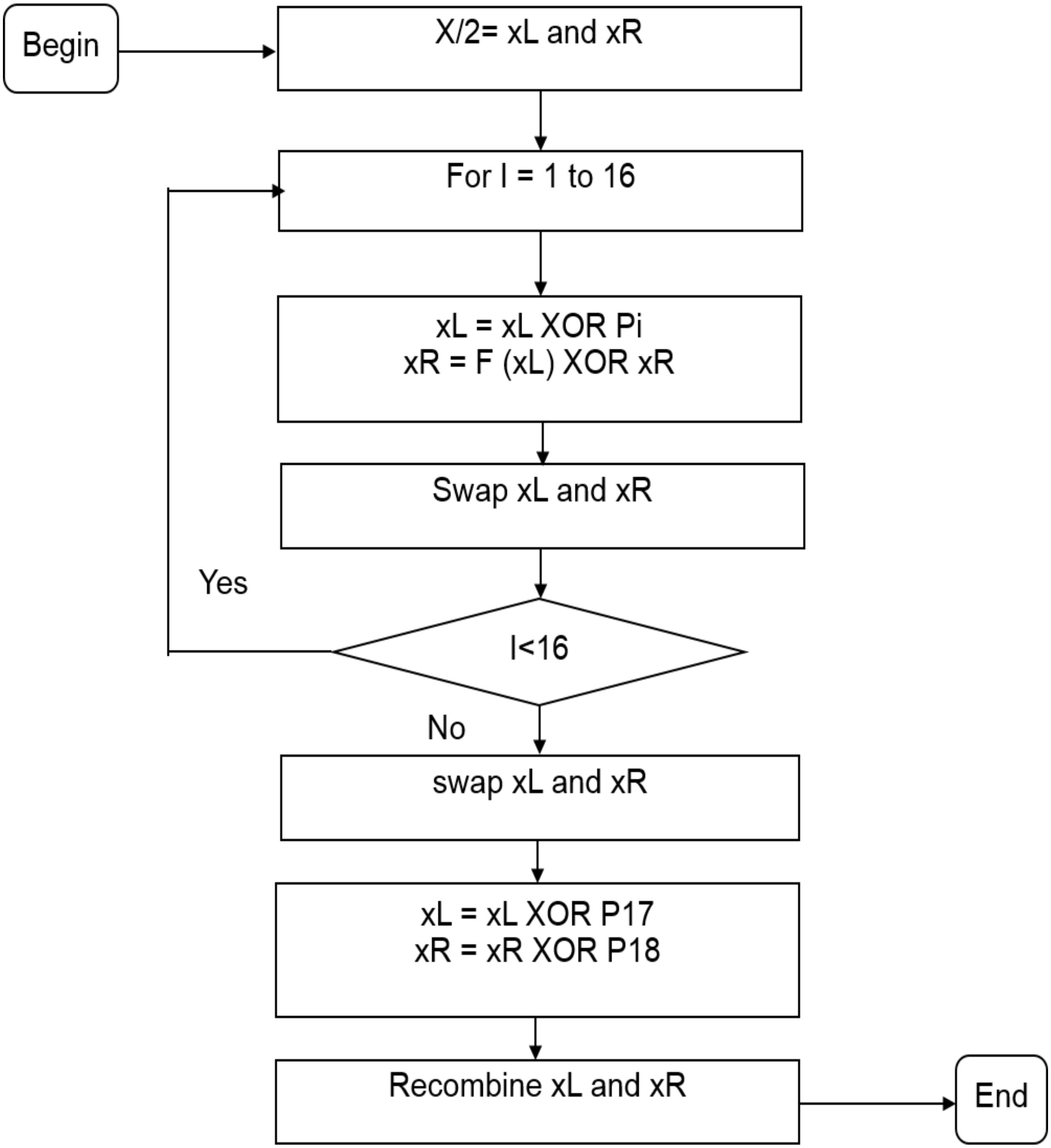


Fig. 7: Blowfish Algorithm

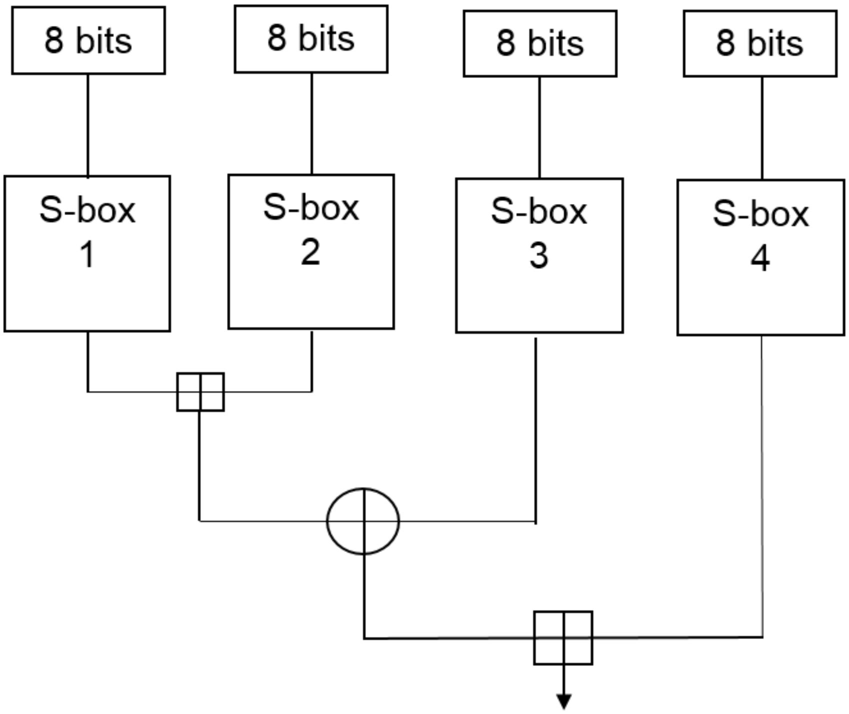


Fig. 8: Function F

2.2 DNA Cryptography

DNA cryptography is used for data encryption as it is fast and store a large amount of data in single DNA strands. The research in this area are going quick and main objective of the DNA computing is to make a machine that will use DNA module for computation and replace silicon-based computer that are we using for processing now a days.

In 1994, Adleman [6] show his work on DNA computing that DNA can be used to solve the complex problem which takes large time when solving using silicon-based computer. He also concluded that DNA has computation latent. His work on this area gives other research to work in Biocomputing. .Normally when we say about DNA computing than we have some knowledge of computation, chemical reaction, and biology. The researcher shows that DNA is like computation machine.

The structure of Deoxyribo nucleic acid (DNA) is discovered by Watson in 1953, shows that DNA structure is the double helix. The first computer based on DNA computing was developed with the concept of “DNA strands are useful to encode an information”. He uses molecular biology laboratory to solve complex problem, in which experiment, he did using PCR (Polymerase Chain Reaction) sequencing, hybridization and gel separation by using test tube in which he hold DNA sequences to perform operation of DNA sequences.

2.2.1 Structure of DNA

DNA stands for Deoxyribo nucleic acid are polymers constructed from monomers called nucleotides. These have a simple structure, made of three components: base, sugar and phosphate. It is the phenomena from which our qualities are defined. It has the power to perform calculations many times greater than any other computation machine laying in this world. RNA and proteins are constructed by DNA contains the genetic instruction needed to construct these RNA and protein. The perplexing structure of the living body comprises of human parts which are the aftereffect of applying straightforward operations to the beginning data encoded in a DNA arrangement called qualities. Similarly, the entangled scientific operation is comprised of straightforward expansion and subtraction. The significant favorable position of DNA which consists of four bases, Cytosine(C), Guanine (G), Thiamine (T) and Adenine (A).

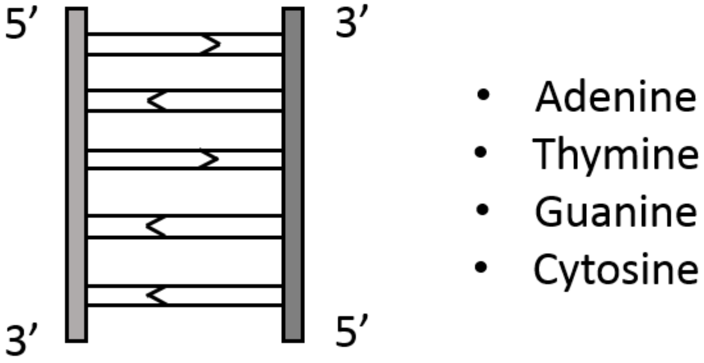


Fig. 9: Four Nitrogen Bases

Watson Crick [7] shows the DNA complementary condition. According to this condition, in double helix DNA, DNA strands each have four bases bind together using hydrogen bond in pairs as A is bind with T using two bonds and C is bind with G using three bonds, which makes the double helix structure of DNA strands. The complementary property as A bind with T and C is bind with T makes it a unique data structure, which will be used to solve the complex problem of mathematics. It uses the chemical reaction such as Hybridization and Litigation to solve the problems. In hybridization chemical reaction, two single strands DNA sequences are combined together in such a way that every base is connected to its complementary base, not any other base makes a new single double strand DNA. In litigation chemical chain reaction, the two double strand DNA modules are bind together to make a single new double strand DNA module. The DNA bases can be encoded using digital encoding design a computer which is bio-computer as it used the DNA bases for computation, which will replace silicon-based computer in the near future as they are faster than the silicon-based computer. DNA have some issue as they are damaged by UV energy comes from the sun and from thermal energy. DNA enzymes are used to when some problem occur in DNA strands like when one of the DNA strands of double helix structure is damaged then enzymes is used to restore the DNA strands using DNA complementary rules. When it is not able to restore the DNA strands means it is enabled to insert T for its complementary A then simple it cut that DNA strands. DNA double helix structure make it use it used in many applications. As the DNA sequence is large no one not knows which DNA sequence is used, so it is used as a one-time pad for many applications which uses the one-time pad. In DNA, we perform the mathematical operation as addition, subtraction, xor etc., between four bases. The result of each operation gives one the base.

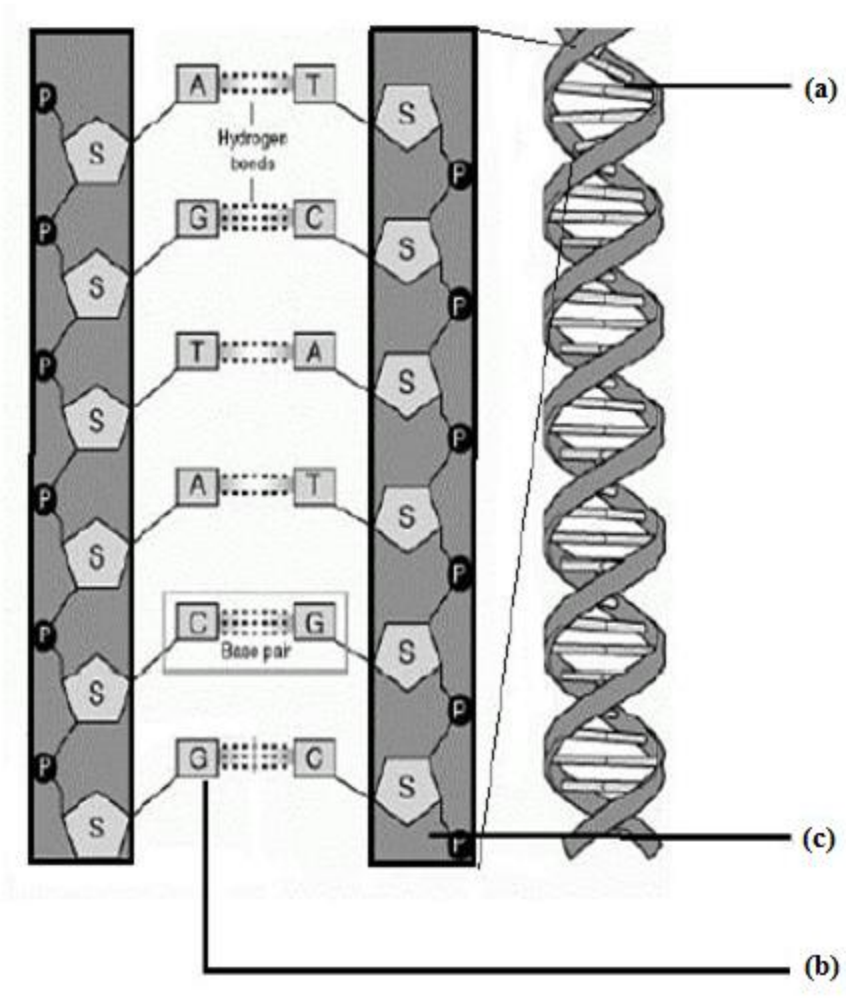


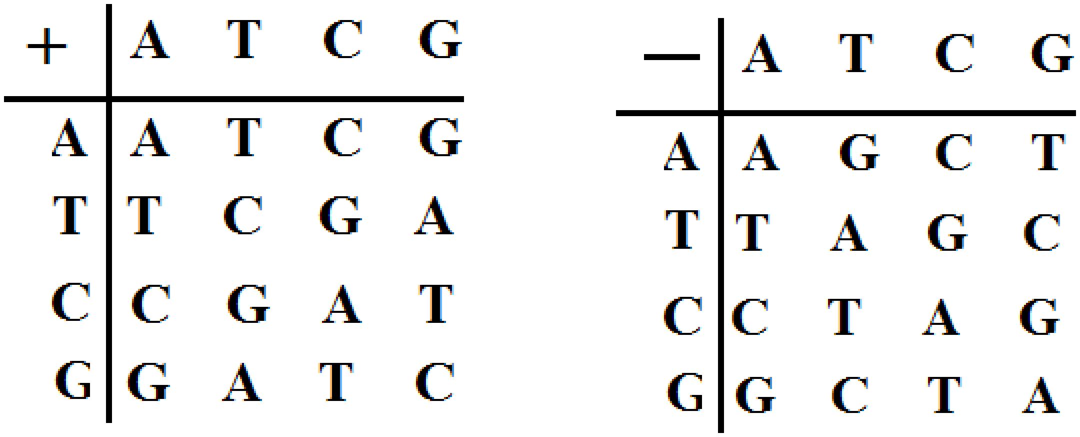
Fig. 10: Molecular Structure of DNA (a) Double Helix Structure of DNA (b) Base pair of DNA (c) Sugar Phosphate Backbone

2.2.2 Addition and Subtraction operation on DNA sequences:

With the rapid developments of DNA computing, some biology operations and algebraic operations based on DNA sequence are presented by researchers, such as addition operation. Addition and subtraction operations for DNA sequences are performed according to traditional addition and subtraction in mathematics. For example, 11 + 10 = 01, 01-11 = 10.

We use 00, 11 to A & G and 01, 10 to T & C because A & G and T & C are the complements to each other. The detail description of addition and subtraction rules of DNA sequences is shown in

Table 1 and Table 2. In this thesis, we will use this wonderful addition rules to scramble the pixel value of the original image.



+	A	T	C	G
A	A	T	C	G
T	T	C	G	A
C	C	G	A	T
G	G	A	T	C

—	A	T	C	G
A	A	G	C	T
T	T	A	G	C
C	C	T	A	G
G	G	C	T	A

Fig. 11: DNA Addition and Subtraction

2.3 MD5 Hash Function

MD5 encryption is used in cryptography to encrypt the data. It is a one-way function means it is not reversible. MD means Message-Digest is a mathematical function, which take a variable length input for processing. Hare, the number 5 show the version number of MD function came after MD4 version.

Message digest 5, which is also known as MD5 gives a 128 bit (16 Byte) hash value is a cryptographic hash function that takes input of variable length. According to experts, the message digest is commonly represented as 32-digit hexadecimal form. MD5 used in security functions which were identified by standard Engineering Task Force (IETF).

MD5 was created by Prof. Ronald L. Rivest of MIT. It is the 3rd version in this md algorithm series. The other two versions of MD series is MD2 MD4. Commonly all version of md series have quite a same structure. Mainly MD2 version of message digest series working on 8-bit computers while the other two version of MD series MD4 and MD5 were work on the 32-bit machine.

MD5 algorithm are used to validate any file that this file belong to a particular person means this algorithm used to authenticate the user.

This algorithm widely used in DS digital signature. In digital signature applications, it is used to compress the large file before encryption using any public key cryptographic algorithm.

This algorithm is used for storing the password in the database. Since MD5 is a one-way function means it cannot be reversed, so the password of the user is safe and secure. No one can know another user password. So today all websites which want to register the user, save their password using a hash function. So when they login using username and password they cannot match password they apply a hash function to password then match with database hash function for that particular user. So even website developers do not know the password of the user.

It also used for fingerprint and unique identification of a file, because any change in the file change the hash value of that file whether the change is small or large. So it is used for checking the integrity of the file. In general MD5 is slower than MD4 as it is more compact and highly coded than MD4, but as per the security, it provide more security than the previous version of MD series.

In any hash function, it is infeasible that two different messages have same Digest, but if two different messages have same message digest than it led to multi-collision attack problem. Basically collision is occurring for many, one-way hash function, but due to heavy deployment of MD5 hash function may sometimes it gives the problem. So when any hash function is designed, then it is main concern that it does have any collision for two different messages.

2.3.1 History of MD5

MD5 is an algorithm developed by Professor Ronald Rivest of MIT University in the series of Message-Digest Algorithms. When statistics and analytics indicated that the predecessor of MD5, i.e. MD4 algorithm is quite insecure and vulnerable, MD5 was designed in 1991 to be more securable and conservative replacement against attacks.

In 1993, Dan Boer and Bosselaers were succeeded partially in finding that an identical message digest is produced by two different initialization vectors.

In 1996, Dobbertin shows a collision of the compression function of MD5 algorithm. This was actually not an attack over the whole MD5 function, but it suggested considering any better cryptographic replacement for use.

On March 18th, 2006, Klima published an algorithm which used only one notebook computer and capable for finding a collision in MD5, using a method particularly known as Tunneling.

In 2008, United States Cyber Command for their mission statement used the MD5 hash as a part of their official emblem.

On December 24th, 2010, Tao Xie and Dengguo Feng announced the first published MD5 collision in single-block (two 64-byte messages which have same MD5 hash). Previous collision discoveries relied on multi-block attacks. Xie and Fend, for some reasons, didn’t disclose the new attack method. They have challenged the Cryptographic community of $10,000 for the one who finds any other 64-byte collision method before January 1st, 2013.

2.3.2 MD5 Algorithm

The MD5 hash function is an extension of MD4 hash function. It is a chain digest algorithm that take a data block of 512 bits at a time which is organized as little endian. At a time, one block is processed. Hare we provide the initial seed to generate the first digest after the process with the first block of data. So after this processing we have digest that is used as a seed for next data block. After last data block is processed, its digest is used as the final digest for the entire stream. That processing prohibits the use of parallel processing as the digest of the previous step is used as the seed for next step. The whole is repeated for four times and each round it uses one of the four nonlinear functions. After we get final 128-bit digest.

Let us suppose that we are giving an input message of b-bit and we want to find its message digest. Here, b is an arbitrary integer which can be equal to greater than zero but no less than zero.

Therefore, let the bits of the input message be as follows:

m0 m1 m2 ... m (b-1)

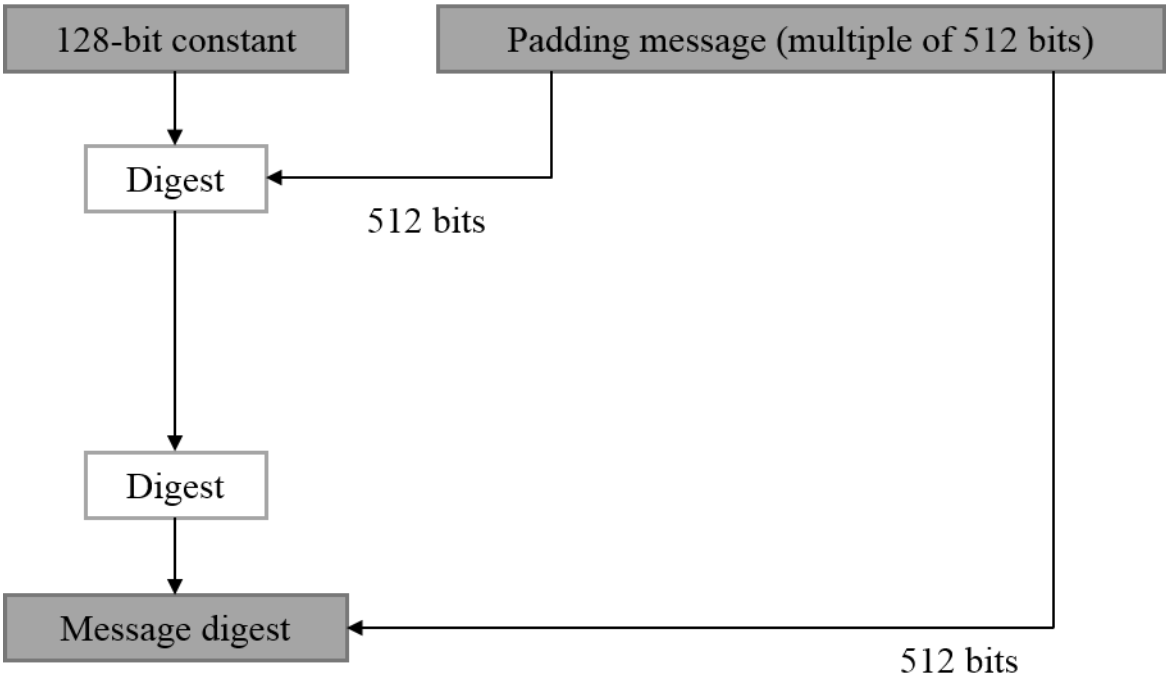


Fig. 12: MD5

The MD5 Algorithm takes in concern the following steps to compute the Message digests of the given input message.

Padding the input message

The b-bit message is extended or more specifically “padded” so that its length reaches 448 bits, which is multiple of 512 bits. The padding is performed by appending a single “1” bit to the input message and rest bits are filled with “0” bits so the length of the block in bits in equal to the 448 bits.

Append 64-bit number

Now, length of the input message is converted to 64-bit representation and then it is appended to the previous result. After this processing, the resulting message of padded bits and b has a length which is multiple of 512 bits. Also, this message is an exact multiple of 16 (32-bit) words. Thus, m [0 ... n-1] denote the resulting message words; where n is a multiple of 16.

Message-Digest Buffer initialization

The message digest of a message is computed using 4-word buffer (A, B, C and D). Here, each of A, B, C and D buffer is a 32-bit register. Each register is initialized with a hexadecimal value which is given as:

Word A: 01 23 45 67

Word B: 89 ab cd ef

Word C: fe dc ba 98

Word D: 76 54 32 10

Processing of each 16-Word Blocks

Firstly, four functions are defined that each function take three input each of size 32 bit and gives an output of one 32 bit word.

They are as follows:

F(X, Y, Z) = (X AND Y) OR ((NOT X) AND Z)

G(X, Y, Z) = (X AND Z) OR (Y AND (NOT Z))

H(X, Y, Z) = X XOR Y XOR Z

I(X, Y, Z) = Y XOR (X OR (NOT Z))

Hare, the 64-element table T [1…..64] is made using sine function. Let ith element of element table is denoted as T[i], then its value is equal to 4294967296 times of abs(sin(i)) integer part. Where I used in radians.

Output

The output A, B, C and D are produced by message digest. In which we begin with A as lower byte and ends with D as a higher byte.

Some optimization limitations are in this algorithm due to some properties of operation which are used in this algorithm as:

- Additions can be reordered by commutative laws.
- Rotate does not distribute over addition.
- Addition does not distribute over rotation or logical.

2.4 Logistic Chaotic map

The logistic map is a type of chaotic map used to generate the values between 0 and 1. It is a one-dimensional map of degree 2 polynomial. Chaotic map is sensitive to the initial value of parameter means changing in the initial value of parameter there may be a huge change in the result that is computed using chaotic map. These systems have dynamic behavior so they are also used to generate the pseudo random values.

The equation of logistic map is given as:

$X_{n+1} = \mu X_n (1-X_n)$

Where:

Xn is a number between [0, 1] and μ has optimal value lie between 3.569945 and 4.

Behavior dependent on μ

The behavior of this system is dependent on the initial parameter μ. When we change the value of μ then system behave in different manner. System behavior on the different value of μ is observed:

When the value of μ between 0 and 1, the population is not dependent to the starting population and population will die means population not generated.

When the value of μ between 1 and 2, the population increase, but it also not dependent to the starting population.

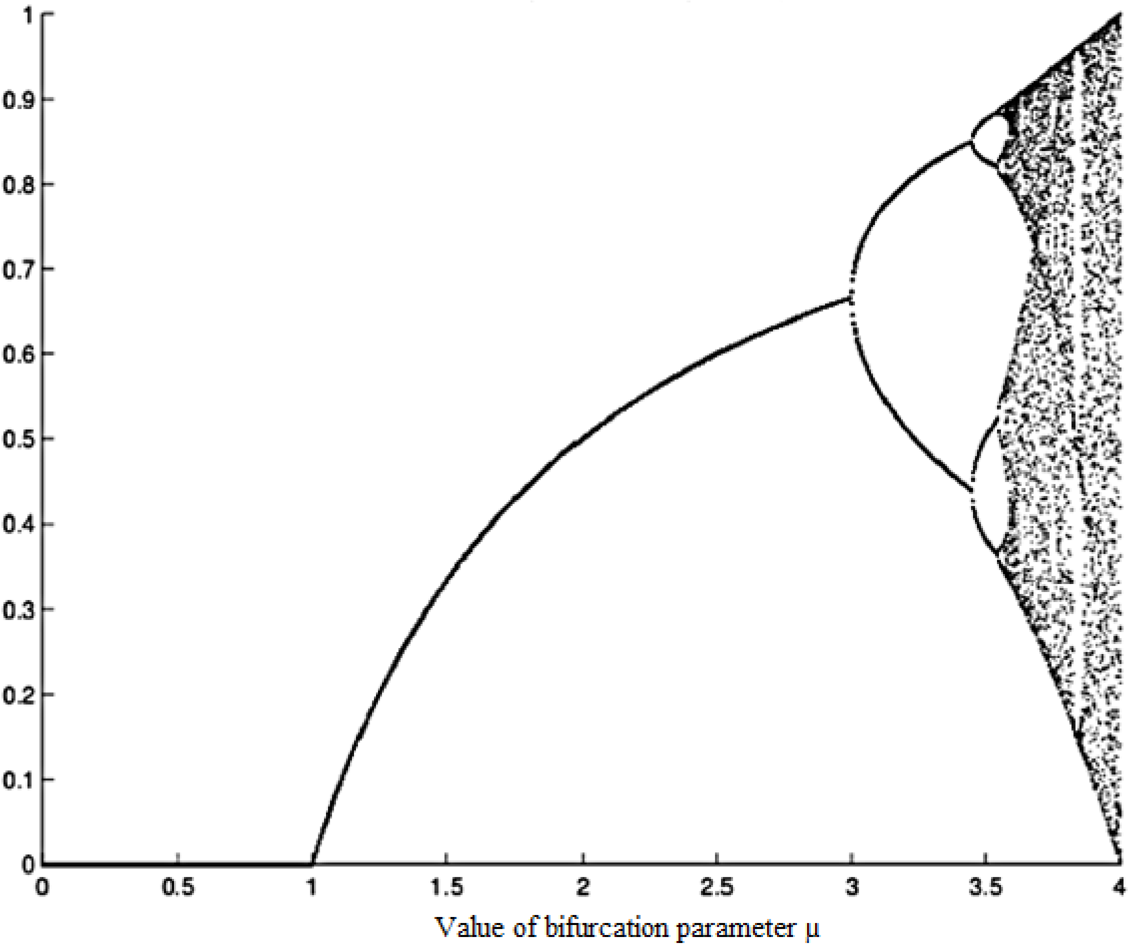


Fig. 13: Bifurcation diagram for logistic map

When the value of μ between 2 and 3, the population increase, but it will fluctuate to that value for some time and atμ =3 it slow down.

When the value of μ between 3 and 3.44949, the population will oscillate between two value and these value are dependent on initial parameter μ. The values are same for all initial conditions.

When the value of μ between 3.44949 and 3.54409, the population will oscillate between two value and these value are dependent on initial parameter μ.The values are same for all initial conditions.

When the value of μ is greater than 3.54409, the population will oscillate among 8 then 16, 32, 64 values. Every time when values of oscillations increase the length of oscillation decreases.

The value of μ greater than 4, values leave the interval [0, 1] for all initial values.

So the value of μ = [3.56995, 4] is best for the logistic chaotic map.

Chapter 3

Literature Survey

Color Image Encryption Algorithm using DNA Code and Chaos Theory

In this paper, the author used a combination of DNA computing and chaos theory for image encryption. The chaotic map is used to scrambling the image pixel and DNA computing is used to change the value of image pixel. Hare DNA sequence is used as a one-time pad (OTP). It used the 1-dimensional chaotic map to change the location of image pixel [9].

Firstly the original image is taken and separated into three color component and also enter the starting value for the 1-dimensional chaotic map then it will generate the X1, which is dependent on original image pixel value. After this three chaotic sequences are created for three different component of the image. Now change the image pixel value by XOR the image pixel value with the chaotic sequence for three color component, and get 3-dimensional image matrix as c1.

Now change the image pixel location using following equations:

C2 (i, j) =C1 (lnx1 (i), lnx2 (j))

C3 (i, j) =C2 (lnx1 (i), lnx2 (j))

Where lnx1, lnx2 are the matrix of string x1, x2.

Now again change the pixel value using DNA computing. Taking a DNA sequence as one time pad from large number of DNA sequence and then XOR this with C3 DNA matrix and getting C4 matrix of three color component, then convert this DNA matrix to decimal matrix then to image and we get encrypted image. The reverse of this process we get our original image.

Image Encryption Algorithm based on DNA Biological Properties and Chaotic Systems

In this paper, the author combines the biological properties of DNA cryptography and chaotic sequence to encrypt the image. DNA cryptography is used to scramble image pixel and chaotic sequence is used to change the pixel value [10].

Firstly the image matrix change to binary then based on conversion table binary values change to DNA encoding. Now complement each DNA sequence as like ‘ATGC’ to ‘GCAT’. And also convert the DNA sequence to PCR DNA sequence. After this convert these DNA sequence to the binary then decimal sequence. After this 8-bit XOR these sequence and get single matrix. Now rotate the original matrix by 180* clockwise rotation and XOR with the previous result. After this, we use improved chaotic sequence to change the value of the pixel. Then XOR the result with the previous result and we get encrypted image.

Improved chaotic sequence equation as follows:

$S^{’} = (S * C) \bmod 256$

S is the original 1-dimension chaotic sequence and $S^{’}$ is the improved chaotic sequence, C is constant (C = 100000000).

A Composite Image Cipher using DNA Sequence and Genetic Algorithm

In this paper, the author used DNA-based image encryption and genetic algorithm to encrypt the image. Hare genetic algorithm is used to find out best DNA mask. To find out, best DNA mask genetic algorithm repeat several times. Hare author use 8 complementary DNA rules [8].

Firstly a 120-bit key is taken. The initial value of chaotic sequence is calculated as

$X_0 = (k1 \text{ xor } k2 \text{ xor } k3 \text{ xor } \ldots\ldots. K8 + \sum k_i)/2^{12}$

The initial population is created using chaotic sequence. Initially an empty mask is created of dimension M X N and value of the mask is calculated using the equation as $P_i = X_i \text{ x } 255$. Hare some number of such populations are created to generate a cipher image. Now the original image and every mask is converted into 1-dimansional DNA sequence using 8 complementary rules. Rule number of encoding each value is given by $R_i = (X_i \times 7) + 1$. Now XOR operation is performed between the original image and each DNA mask. After this, entropy is calculated for each encrypted image. After this genetic algorithm is applied. The entropy value is used as a fitness function in the genetic algorithm. A single point crossover is used for every pair to permute the sequence and generate the new two child sequences. At each iteration, that solution has entropy value less than the minimum entropy value are replaced by new random solutions. This step is continued until we get entropy value greater than or equal to defined entropy value or after a fixed number of the round. At the end, a solution which has the highest entropy is chosen as a cipher image.

A Multilevel Image Encryption Algorithm Based on Chaos and DNA Coding

In this paper, the author used the combination of DNA Coding and Chaotic map to encrypt the image. DNA coding is used to change the value of image pixel and the chaotic map is used to scramble the image pixels. Hare chaotic map also used to generate the DNA sequence [11].

Firstly an image is taken and converted to binary then finally in DNA sequence matrix. As each decimal value convert to four DNA bases we divide the DNA matrix to four sub matrixes. Now generate the four different chaotic sequences using four different initial value. Also, we scramble the four DNA sub-matrix based on the index value of chaotic sequences. Now convert each chaotic sequence into DNA chaotic sequence using DNA rules. Now we perform addition operation between DNA sub-matrix and chaotic DNA matrix and get the resultant matrix. Now join all sub-matrix and get final DNA sequence matrix and then convert this DNA matrix to decimal and we get cipher image. The original image is getting by reverse the whole process.

Hybrid Approach of Image Encryption Using DNA Cryptography and TF Hill Cipher Algorithm

RD hill cipher algorithm is the combination of DNA cryptography and TF hill cipher. In this algorithm, firstly the image is converted into DNA matrix and then DNA to amino acid and then hill cipher algorithm is applied to amino acids [12].

Firstly an image is taken and converted into a binary matrix. Now replace the first half of nibble with the second half of nibble. And then convert it into DNA sequence matrix. Now convert the DNA sequence to the amino acid. For converting the DNA sequence into the amino acid, we divide DNA sequence into three letters group. Each three letter group is called codons. Since 4 bases in letter combinations, we have total 64 combinations. In these 64 codons, some are stop and nonsense codons. There is total 20 Amino acid. So some codons may change to same amino acid, so we have to maintain which codon change to which amino acid, so after every amino acid we put ambiguity number to understand. Now hill cipher is applied to this amino acid to get cipher image.

Chapter 4

Proposed Work

4.1 Introduction

Our proposed algorithm is a combination of DNA cryptography and Blowfish encryption. DNA cryptography encrypts data bit level means it takes two bits at a time and encrypt it and blowfish algorithm take a 64-bit block data and encrypt it. So our algorithm encrypts data at the bit level and block level. So our algorithm is basically a double encryption algorithm for generating more secure and efficient cipher text with the help of DNA cryptography and blowfish algorithm. First, it encrypts data at the bit level and it encrypts data in blocks.

4.2 Proposed Algorithm

In our algorithm, we use logistic chaotic map. In the logistic chaotic map, we have to initialize the initial value at the start of the algorithm and it is fixed for the whole algorithm and we never change these values. So we modify this step and the initial value calculated using the blowfish key. And this initial value is calculated using a hash function.

Our encryption algorithm has some steps as follows:

Convert image to DNA matrix

In this step, we take image and key and convert it into the matrix. Our algorithm uses a color image (RGB image) so our matrix has three dimensions. For conversion from image to DNA first we convert our image matrix to binary matrix. After getting a binary matrix, we convert it into a DNA matrix using DNA encoding rules. DNA contains four nucleic acid bases T (thymine), G (guanine), A (adenine), C (cytosine), where A & T are the complements, G & C are complements. In our algorithm, we use 00,01,10,11 to denote C, A, T, G respectively. So for 8-bit image each pixel can be expressed a DNA sequence of length 4.

For example: in the original image, the value of the first pixel in binary form is 01100010 then its DNA sequence is ‘ATCT’. So we convert the binary matrix to DNA matrix using these DNA encoding rules.

Generating chaotic sequence matrix and convert it to DNA chaotic matrix

In this step, we are generating a chaotic sequence and convert it into the DNA chaotic matrix. We use one-dimensional logistic map for the generation of the chaotic sequence. In logistic map initial value of ‘x’ is fixed and never changes with any number of image encryption processes. So it may be possible to crack the initial value of ‘x’ by some hacker and it will miss use it.

So in our algorithm we initialize the value of ‘x’ based on blowfish key. So an initial value of ‘x’ is not fixed and its value depends on the input key. So its value varies in the different image encryption process. For the generation of the initial value of ‘x’, we pass key to the MD5 hash function. The Md5 hash function gives 16, the 32-bit value. So we add all these values in mod 2^{16} and divided it by 216 to get an initial value of ‘x’ which lying between 0 and 1.

Now using this logistic map we generate a chaotic sequence matrix and then convert it into the binary matrix and then convert it into the chaotic DNA sequence matrix.

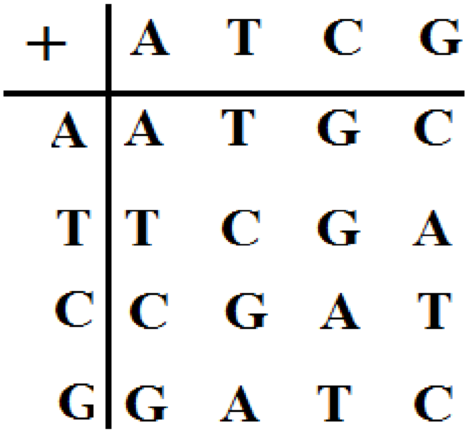
DNA sequence addition

In this step, we perform the addition of image DNA sequence matrix with chaotic DNA matrix and also perform the 90* anti-clockwise rotation.

First, we divide the image DNA matrix into a 3x3 matrix and rotate each 3x3 matrix 90^* anti-clockwise. Then we again merge all 3x3 matrices to DNA matrix.

Now we perform the DNA addition between image DNA matrix and chaotic DNA matrix [13].

After the addition operation, we get encrypted matrix. Here we encrypt data at the bit level. Now we convert DNA matrix to the binary matrix, then we convert back to the decimal matrix and get a new sequence matrix.



+	A	T	C	G
A	A	T	G	C
T	T	C	G	A
C	C	G	A	T
G	G	A	T	C

Fig. 14: DNA Addition

Matrix rotation

In this step, we rotate matrix 90^* anti-clockwise direction. First, we divide the matrix into 3x3 block matrixes and then rotate each 3x3 block matrix 90^* anti-clockwise direction and we get scrambled image. Hare we rotate the decimal value in anti-clockwise direction and in DNA sequence addition step we rotate bit value in an anti-clockwise direction. After rotation, we combine these 3x3 matrix blocks.

Blowfish encryption

In this step, we encrypt our matrix data using blowfish algorithm. Blowfish algorithm takes 64-bit data (8 bytes) for encryption at a time and we get 64-bit encrypted data. So we divide our matrix into a 64-bit data block and then feed into blowfish algorithm and also key which is provided by the user.

After blowfish encryption, we get our cipher data and recombine into a single matrix and after converting back into the image and we get our cipher image. The size and dimension of our cipher image are same as the original image. There is no change in size and dimension of the image.

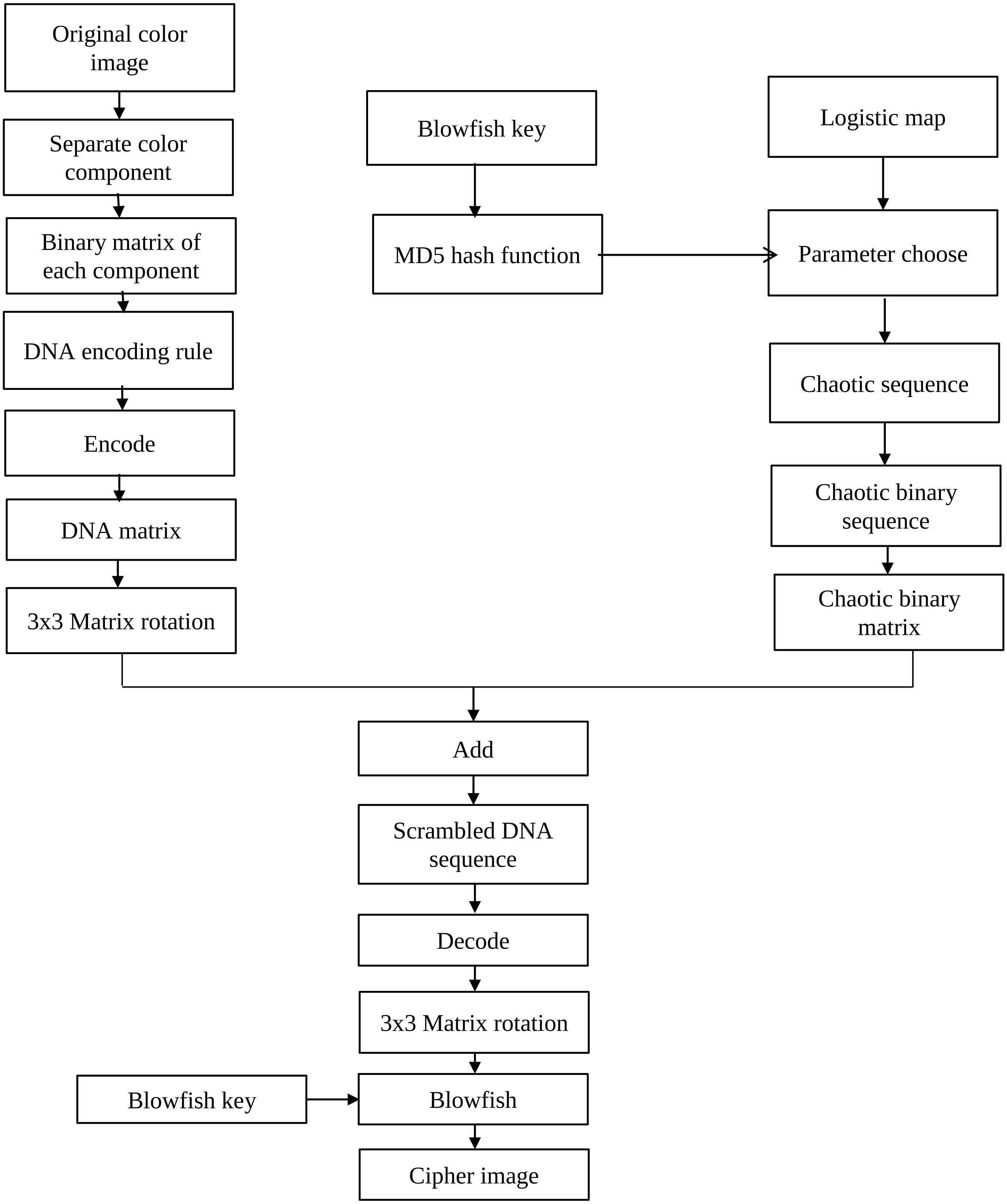


Fig. 15: Proposed Algorithm

Chapter 5

Experiment and Result Analysis

5.1 Overview

With the application of an encryption algorithm to an image, the values of image pixel changes when original image is compared with the encrypted one. Those changes should be made in an irregular manner by a good encryption algorithm and also the difference in pixel value difference between the original image and the encrypted images are also maximized. Also, to get a good encrypted image. It must be composed of totally random patterns that don't reveal any of the features of the original image. The encrypted image has to be independent of the original image. The correlation should be low with the original image.

5.2 The Gray Histogram Analysis

One desired property of a block ciphers is that cipher-images histogram is uniformly distributed so that the cipher-images do not be arany statistical information about their corresponding plain-image. An image histogram show how pixels in an image are distributed by graphing the number of pixel at each gray level. Histogram analysis is widely used to measure the image encryption quality of a block cipher. To prevent the leakage of information to attackers, it is important to ensure that encrypted and original image don't have any statistical similarities. We select an image (512 * 512) and use different key for this same image and we calculate their histogram. Results of histogram comparison between original and encrypted image are shown in Fig. 15 for different blowfish key.

To analyze the statistical performance we compare the gray histogram of the image before and after encryption. Fig. 16 shows the gray histogram of the original image and encrypted image. From the figure, we can see that the original pixel gray values are scattering but pixel gray values after the encryption are concentrated on some value. Clearly it is difficult to use the statistical performance of the pixel gray value to recover the original image.

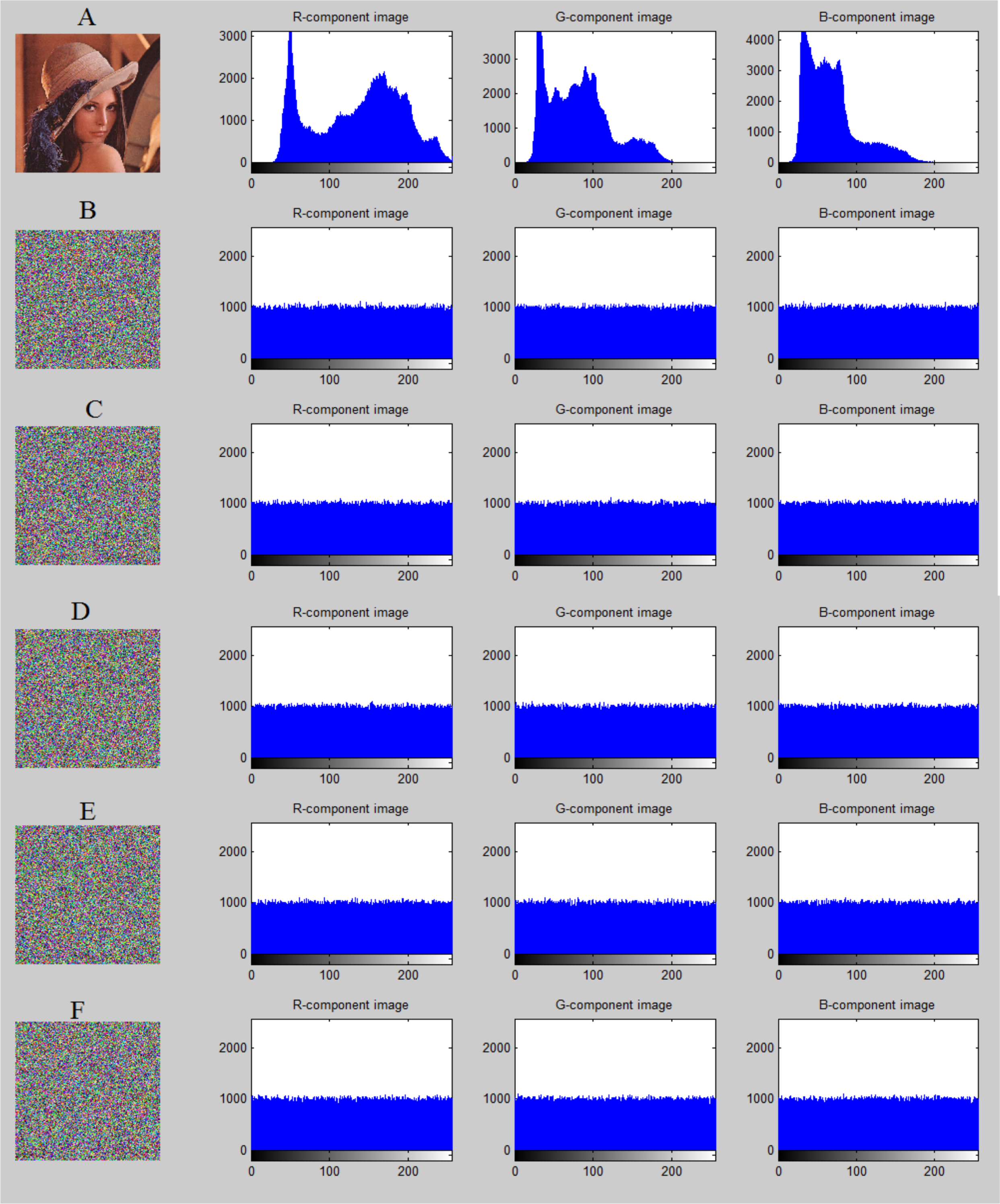


Fig. 16: Result of histogram comparison encryption by different key. (A) Original Image (B) Encrypted image using “testkey” (C) using “encrypt” (D) using “computer” (E) using “hello” (F) using “Lenovo”

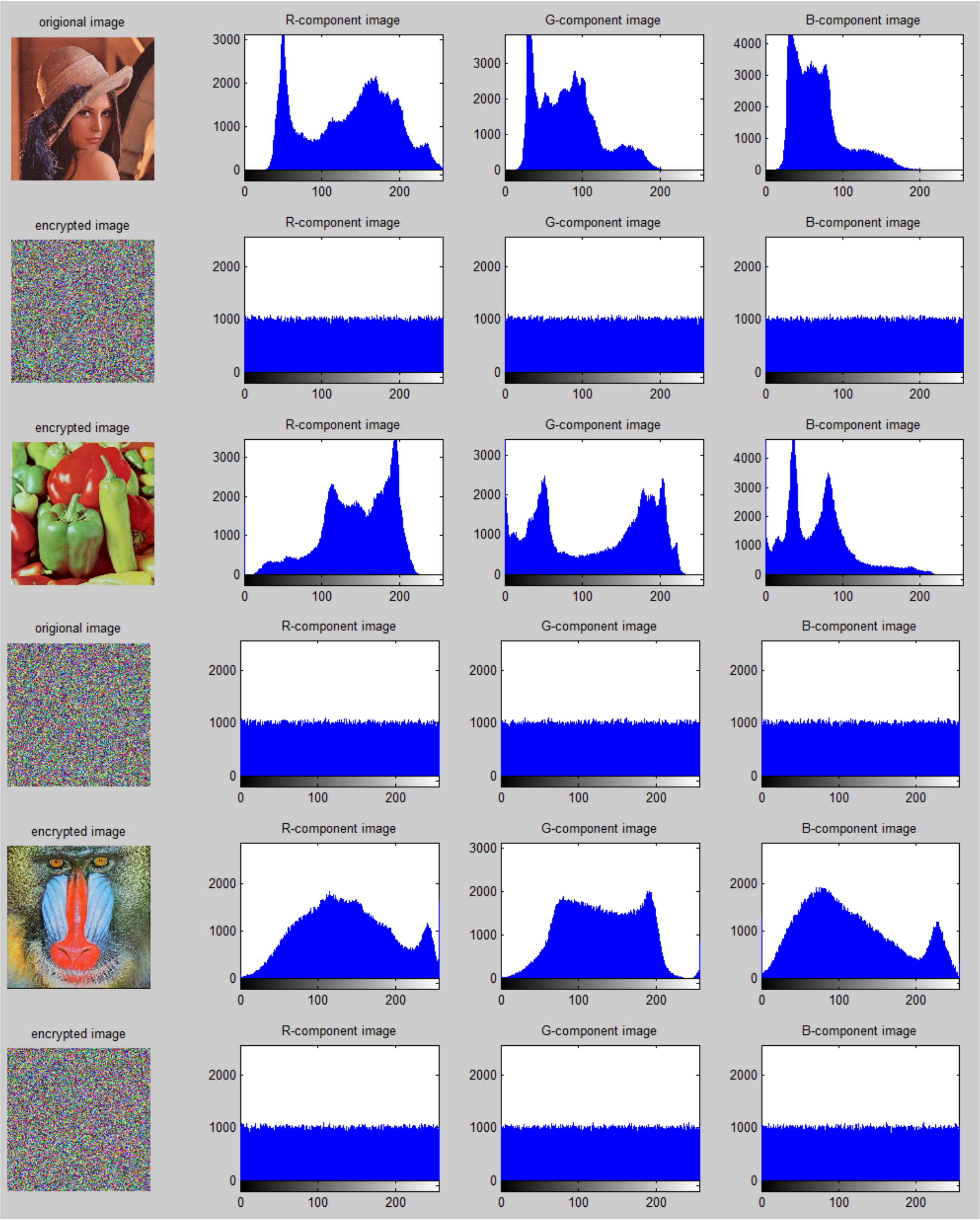


Fig. 17: Comparison of histogram between Original image and Encrypted image

5.3 Correlation Coefficient Analysis

Correlation determines the relationship between to variables. In other words, correlation is a measure that computes the degree of similarity between two variables. The correlation coefficient is a useful measure to judge the encryption quality of any cryptosystem [40]. Any image cryptosystem is said to be good if encryption algorithm hides all attributes of a plaintext image, and encrypted image is totally random and highly uncorrelated. If encrypted image and plaintext image are completely different then their corresponding correlation coefficient must be very low or very close to zero. If the correlation coefficient is equal to one, then two images are identical and they are in perfect correlation. In case of perfect correlation (correlation coefficient equal to 1), encryption process completely fails because the encrypted image is same as the plaintext image. When the correlation coefficient is -1 then encrypted image is negative of original (plaintext) image. In short, the correlation coefficient between an image and itself is 1, the correlation coefficient between an image and totally uncorrelated image is zero and correlation coefficient between an image and its negative is -1. Let x and y be the gray scale value of two pixels in the same place in the plaintext and ciphertext images. Then mathematically correlation coefficient can written as:

$$C.C = \frac{Cov(x,y)}{\sigma x \times \ \sigma y}$$

$$\sigma x = \sqrt{VAR(x)}$$

$$\sigma y = \sqrt{VAR(y)}$$

$$VAR(X) = \frac{1}{N}\sum_{i=1}^{N}(x_i - E(x))^2$$

$$Cov(x,y) = \frac{1}{N}\sum_{i=1}^{N}(x_i - E(x))(y_i - E(y))$$

Where C.C is correlation coefficient, Cov(x, y) is covariance at pixels x and y, x and y are gray-scale values of plaintext and ciphertext images, VAR(x) is variance at pixel value x in the σ_x

plaintext image, is standard devotion, E is the expected value operator, N is the total number of pixels for N*N matrix.

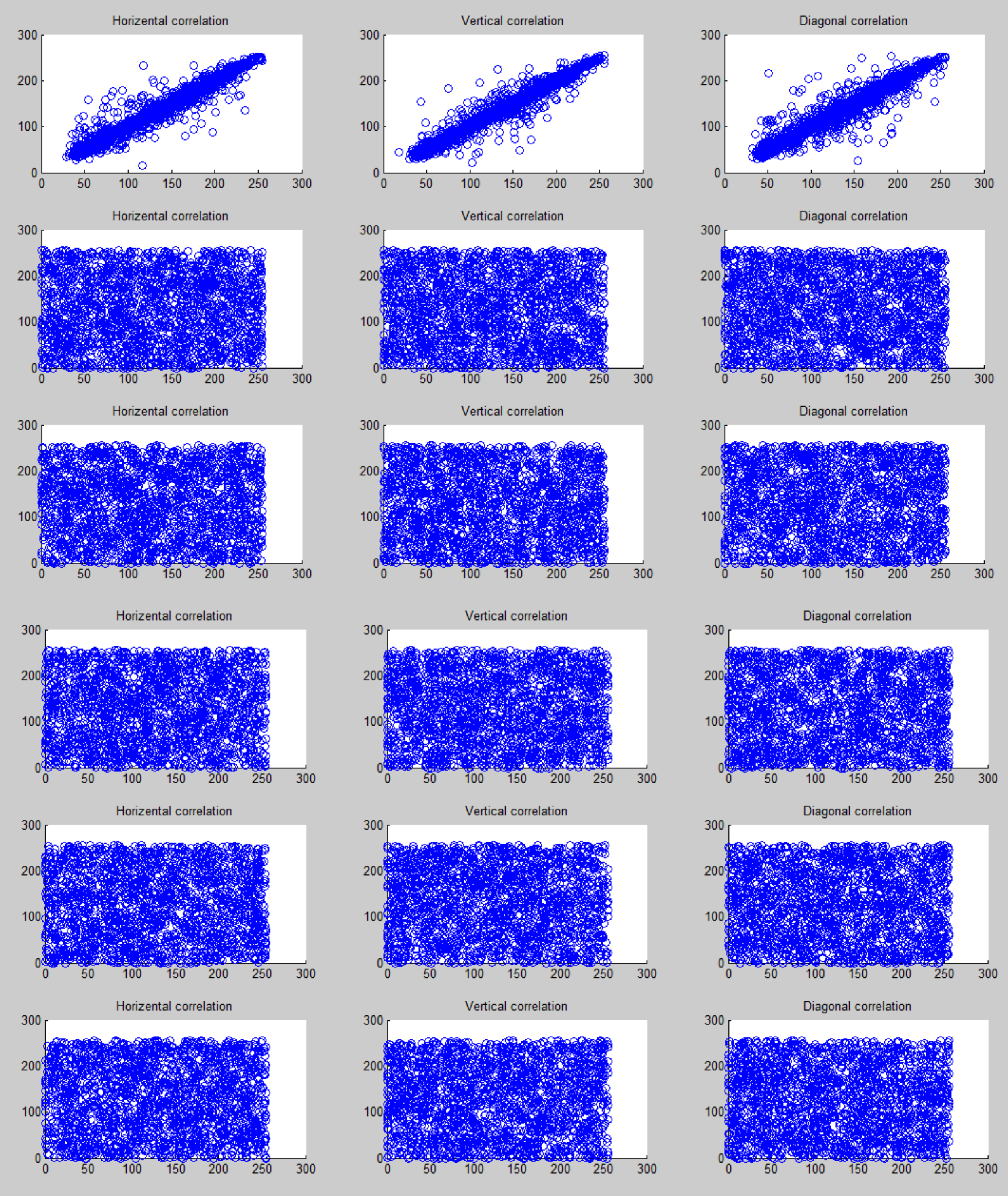


Fig. 18: Correlation Comparison of Original image and Encrypted image for different key on same image

Table 1 Correlation Coefficient of value of original image and encrypted image

Lena	Pepper	Monkey			
0.9749	0.9620	0.8923	Horizontal	Original Image	Color Component Red
0.9782	0.9572	0.8338	Vertical		
0.9593	0.9230	0.7903	Diagonal		
0.0004	0.0029	0.0019	Horizontal	Encrypted Image	
0.0002	-0.0014	-0.0026	Vertical		
-0.0020	-0.0009	0.0018	Diagonal		
0.9632	0.9840	0.8446	Horizontal	Original Image	Color Component Green
0.9729	0.9815	0.7641	Vertical		
0.9497	0.9676	0.6980	Diagonal		
0.0008	0.0004	-0.0005	Horizontal	Encrypted Image	
0.0001	0.0007	0.0014	Vertical		
-0.0016	-0.0031	0.0001	Diagonal		
0.9376	0.9644	0.9046	Horizontal	Original Image	Color Component Blue
0.9515	0.9699	0.8567	Vertical		
0.9212	0.9460	0.8153	Diagonal		
0.0006	0.0007	-0.0020	Horizontal	Encrypted Image	
-0.0004	0.0021	-0.0014	Vertical		
0.0016	0.0010	-0.0029	Diagonal		

5.4 Information Entropy

Distribution of gray value in the image is measured using information entropy. Higher the information entropy value, more uniform distribution of gray value in the image and the lesser value of information entropy give the non-uniform distribution of the gray value of image pixel. Information entropy is the main feature of uncertainty. It shows the degree of uncertainties in any communication system the entropy, H (m) of any image can be calculated as:

$$H(m) = \sum_{i=0}^{2^N - 1} p(m_i) \times log_2 \frac{1}{p(m_i)}$$

Where p(m_i) represent the probability of occurrence of the symbol m_i. let us consider a true random source that generate 2^8 symbols with equal probability i.e., m = {m1 … $m_{2}{}^{8}$}, where each symbol is represented by 8 bits. If the equation is evaluated for the aforementioned case, its entropy obtained is H(m)= 8 bits, which corresponds to a uniform random source.

Hare we use Lena 512*512*3 color image to calculate the entropy on the different key. Hare, we also change the length of the key as blowfish also take variable length key.

The original image entropy is: 7.4767

After encryption, we can see that the entropy of cipher image approximately equal to 8, that show how good our encryption algorithm.

Table 2 Entropy of cipher image on different key

Encryption Key	**Entropy**
'testkey'	**7.9998**
'hello'	**7.9998**
'computer'	**7.9997**
'lenovo'	**7.9998**
'encrypt'	**7.9997**

Chapter 6

Conclusion and Future Scope

Conclusion

In this dissertation, we present DNA based cryptography and blowfish algorithm used for image encryption. The proposed algorithm is secure and efficient. DNA cryptography plays a great role in the area of image encryption and cryptography. In these applications, a good encryption algorithm is essential. With extensive simulation studies, it is shown that algorithm proposed in this thesis is one of the secure methods of encryption in the image processing as compared to any other encryption method, because the bit pattern of the message is difficult to crack. As this algorithm uses a variable length key, the key space is much larger than fixed length key. This algorithm has two level of encryption at bit level using DNA cryptography and block level using blowfish cipher so it is much difficult to crack the code. This algorithm work for both gray-scale images and color images. Hare, we use the MD5 hash function for generating the initial value of logistic map so that the value of map not fixed, change as blowfish key change. This algorithm is work with both gray-scale image and color image.

Future work

In this proposed algorithm, we used the combination of blowfish and DNA cryptography to get the good result in the area of image encryption. Hare we use the chaotic map to generate a sequence. The proposed approach could be improved by using a better technique for generation of sequence or could be use different technique for encryption of different type of images.

References

[1] A. Menezes, P. van Oorschot and S. Vanston, “Handbook of Applied Cryptography”, CRC Press, 1997.

[2] H. Delfs and H. Knebl, “Introduction to Cryptography: Principles and Applications”, Second Edition, Springer-Verlag, 2007.

[3] W. Mao, “Modern Cryptography: Theory and Practice”, Prentice Hall, July 25, 2003.

[4] A. S. Tanenbaum, “Computer Networks”, Fourth Edition, Prentice Hall, March 17, 2003.

[5] B. Schneier, “Description ofa new variable-length Key, 64-Bit Block Cipher (Blowfish)”, Fast Software Encryption, Cambridge Security Workshop Proceedings Springer-Verlag, 1994 .

[6] Leonard M. Adleman “Molecular Computation of solution to combinatorial problems” Science, New Series, Vol. 266, No. 5187. pp. 1021-1024 Nov. 11, 1994.

[7] R. J. Lipton,” Using DNA to Solve NP-Complete problems," Science, vol. 268, pp. 542 545, 1995.

[8] Saranya M R, Arun K Mohan “A Composite Image Cipher Using DNA Sequence and Genetic Algorithm”, IEEE International Conference, 2014.

[9] M. Amr Mokhtar, Sameh N. Gobran, EI-Sayad A-M, EI-Badawy, “Colored Image Encryption Algorithm using DNA Code and Chaos Theory”, 5th International Conference on Computer and Communication Engineering, 2014.

[10] Q. Wang, Q. Zhang, X. Wei, “ Image Encryption Algorithm based on DNA Biological Properties and Chaotic Systems”, IEEE International conference, 2007.

[11] Q. Wang, Q Zhang, C. Zhou, “A Multilevel Image Encryption Algorithm Based on Chaos and DNA Coding”, IEEE International Conference, 2009.

[12] R. K. Jangid, N. Mohmmad, A. Dibel, “Hybrid Approach of Image Encryption Using DNA Cryptography and TF Hill Cipher Algorithm”, IEEE International Conference, 2014.

[13] Q. Zhang, L. Guo, X. Xue, X. Wei, “An Image Encryption Algorithm Based on DNA Sequence Addition Operation”, IEEE international Conference, 2009.

[14] S. Bhowmik, S. Acharyya, “Image Cryptography: the Genetic Algorithm Approach”, IEEE International Conference, 2011.

[15] R. Rivest “The MD5 Message-Digest Algorithm”, MIT Laboratory for Computer Science and RSA data Security, 2014