

# **NOVEL CRYPTOGRAPHIC AUTHENTICATION METHOD USING DNA SEQUENCE IN CLOUD ENVIRONMENT**

MAJOR PROJECT SUBMITTED IN PARTIAL FULFILLMENT OF THE  
REQUIREMENTS FOR THE AWARD OF DEGREE OF

Master of Technology

In

Information Systems

Submitted By:

PRIYANSHI JAIN

(2k13/ISY/18)

Under the Guidance

*Of*

N.S. RAGHAVA

ASSOCIATE PROFESSOR



DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

DELHI TECHNOLOGICAL UNIVERSITY

(2013-2015)

# CERTIFICATE

This is to certify that **Priyanshi Jain (2k13/ISY/18)** has carried out the major project titled “**Novel Cryptographic Authentication Method Using DNA Sequence In Cloud Environment**” in partial fulfilment of the requirements for the award of Master of Technology degree in Information Systems by **Delhi Technological University**.

The major project is bonafide piece of work carried out and completed under my supervision and guidance during the academic session 2013-2015. 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.

N.S. Raghava

Associate Professor

Department of Electronics and Communication

Delhi Technological University

Delhi-110042

## **ACKNOWLEDGEMENT**

I take the opportunity to express my sincere gratitude to my project mentor N.S. Raghava, Associate Professor, Department of Electronics and Communication, Delhi Technological University, Delhi, for providing valuable guidance and constant encouragement throughout the project. It is my pleasure to record my sincere thanks to him for his constructive criticism and insight without which the project would not have shaped as it has.

It humbly extends my words of gratitude to other faculty members of Computer Science and Engineering department for providing their valuable help and time whenever it was required.

Priyanshi Jain

Roll No. 2k13/ISY/18

M.Tech (Information Systems)

E-mail: priyanshijain111@gmail.com

## ABSTRACT

---

The rapid evolution of cloud computing services has encouraged many organizations to migrate towards consuming the cloud services, especially towards public cloud services, which is under the control of third party Cloud Services Provider (CSP). Due to which, many security issues gets associated with these services faced by both the cloud service providers and the customers. Thus, the responsibility of resolving these issues goes in both directions i.e. the service provider must ensure a secure infrastructure (the CSP must ensure the data isolation and data segregation) and customer must ensure security of their data through various authentication scheme and cryptographic algorithm. In this work, a novel cryptographic and authentication method is proposed for the cloud environment. The proposed approach ensures both the data integrity from customer's point of view and authentication of enrolled users from cloud service provider point of view. It ensures data integrity by proposing a new symmetric image encryption algorithm using Henon Chaotic systems. Confusion and diffusion is increased by shuffling the image pixel in a specific order for several iterations. This encryption algorithm utilizes the property of the chaotic systems which is known for its randomness and unpredictable behaviour and thus providing a strong cryptographic framework to secure data. The symmetric key (seed) used for encryption is generated through the proposed key generation algorithm based on system entropy information and using DNA science. The ability of DNA to provide large data storage, genetic coding, massive parallel computation and generation of faked and random DNA sequence has increased the possibility of using DNA and gives a prominent direction in cryptographic research. When a user wants to access this encoded data, it needs to get authorized by the authentication server of the cloud. Thus, the proposed user authentication framework enables authentication server to validate the customer request and performs mutual authentication. The statistical analysis on key sensitivity and measurement of image encryption quality proves that the proposed image encryption approach gives a new dimension for secure image transfer. The security analysis, randomness testing of the key using NIST Test suites and resistance to various security attacks like replay attack, man-in-the middle attack proves that proposed scheme serves the objective of the research and helps in lowering down the cloud-computing security concerns like data isolation, data integrity, authentication and account management.

# Table of Contents

<b>Title</b>	<b>Page no.</b>
CERTIFICATE	ii
ACKNOWLEDGEMENT	iii
ABSTRACT	iv
List of Figures	viii
List of Tables	ix
<b>Chapter 1</b>	
<b>CLOUD SECURITY .....</b>	<b>1</b>
1.1 Introduction	1
1.2 Cloud Security Architecture	3
1.3 Need for Authentication in Cloud	4
1.4 Related Work	6
<b>Chapter 2</b>	
<b>DNA CRYPTOGRAPHY AND SEQUENCE ALIGNMENT.....</b>	<b>8</b>
2.1 Introduction	8
2.2 DNA Cryptography	8
2.2.1 Biological Background of DNA	8
2.2.2 DNA Sequencing	10
2.3 Sequence Alignment	13
2.3.1 Sequence Similarity versus Sequence Identity	15
2.4 Dynamic Programming Method	17
2.4.1 Dynamic Programming Method for Global Alignment	18
2.4.2 Dynamic Programming Method for Global Alignment	19
<b>Chapter 3</b>	
<b>CHAOS AND CRYPTOGRAPHY.....</b>	<b>21</b>
3.1 Introduction	21

3.2 Chaos Theory	22
3.3 Chaos-Based Cryptography	23
3.4 Chaotic Map	25
3.4.1 One-Dimensional Map	25
3.4.2 Two-Dimensional Map	26
3.5 Henon Map	27
3.6 Attractor	29
3.7 Confusion and Diffusion	30
<b>Chapter 4</b>	
<b>PROPOSED METHODOLOGY.....</b>	<b>31</b>
4.1 Key Generation Method	32
4.2 Image Encryption Using Henon Chaotic Map	34
4.3 Authentication of User	42
<b>Chapter 5</b>	
<b>EXPERIMENTAL RESULTS AND SECURITY ANALYSIS.....</b>	<b>45</b>
5.1 Key Generation Method	45
5.1.1 Global Alignment and Winning Path	46
5.1.2 Local Alignment and Winning Path	47
5.2 Encryption and Decryption of the Image	49
5.3 Authentication of User	52
5.4 Statistical Analysis	54
5.4.1 Randomness Test	54
5.4.2 Key Sensitivity Test	57
5.4.3 Histogram Analysis	58
5.4.4 Information Entropy Analysis	60
5.5 Security Analysis	61
5.5.1 Man-in the Middle Attack	61
5.5.1 Brute Force Attack	62

5.5.1 Replay Attack	62
---------------------	----

**Chapter 6**

<b>CONCLUSION AND FUTURE WORK.....</b>	<b>63</b>
--	-----------

6.1 Conclusion	63
----------------	----

6.2 Future Work	63
-----------------	----

<b>REFERENCES.....</b>	<b>65</b>
------------------------	-----------

## Figures

<b>Figure</b>	<b>Title</b>	<b>Page no.</b>
Figure 1.1	Cloud Computing Deployment Models	2
Figure 2.1	Structure of Purines and Pyrimidine	10
Figure 2.2	Series of codon in mRNA molecule	11
Figure 2.3	Three zones of nucleotide sequence alignments	14
Figure 2.4	DNA sequence alignment	16
Figure 3.1	Relationship between chaotic system and cryptographic algorithm	24
Figure 3.2	2D representation of Henon map	27
Figure 3.3	The strange attractor	28
Figure 3.4	Attractor for dynamic system	29
Figure 4.1	Flow Chart of key generation method	32
Figure 4.2	Flow Chart for image encryption algorithm	34
Figure 4.3	Padding and Shuffling of image pixel	38
Figure 4.4	Encryption with Byte sequence	40
Figure 4.5	Decryption process	41
Figure 4.6	Flow chart for user authentication and key generation algorithm	43
Figure 5.1	Seed value from DNA	45
Figure 5.2	Global alignment and winning path	46
Figure 5.3	Local alignment and winning path	47
Figure 5.4	Random walk for X value	48
Figure 5.5	Random walk for Y value	48
Figure 5.6	Plot of Henon map	49
Figure 5.7	Encryption process	50
Figure 5.8	Decryption process	51
Figure 5.9	Key exchange and user authentication	52
Figure 5.10	Key sensitivity test	57
Figure 5.11	Histogram analysis	58
Figure 5.12	Man in the middle attack	61



## Tables

<b>Table</b>	<b>Title</b>	<b>Page no.</b>
Table 1	DNA Nucleotide base	9
Table 2	DNA base coding	33
Table 3	Results of NIST Test Suite	56
Table 4	Information Entropy Analysis	60

### 1.1 INTRODUCTION

The Cloud computing technology continues to gain acceptance and considered as a widely used services. It became integral constituent in delivering Information Technology facilities. Every Information Technology houses needs an efficient Information practice and cloud offers significant and efficient services compared to existing traditional on-premises data centre services.

In an October 2009, Peter Mell and Tim Grance of the National Institute of Standards and Technology (NIST) Information Technology Laboratory presentation titled “Effectively and Securely Using the Cloud Computing Paradigm,” presented the definition of Cloud Computing as follows:

***“Cloud computing is a model for enabling convenient, on-demand network access to a shared pool of configurable and reliable computing resources (e.g., networks, servers, storage, applications, services) that can be rapidly provisioned and released with minimal consumer management effort or service provider interaction.”***

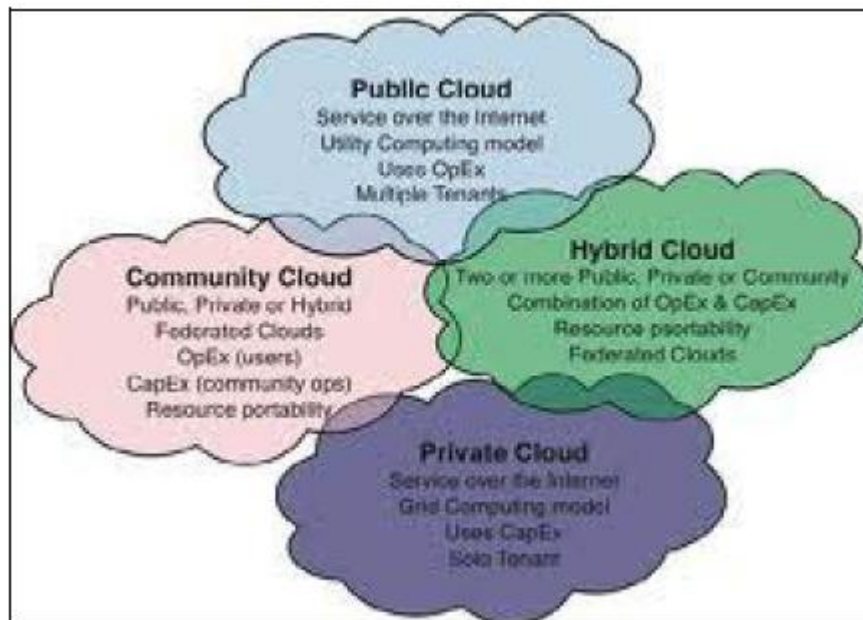
Cloud provides three service model and four deployment model

The models of the services provided by the cloud are described below:

- Cloud Software as a Service (SaaS)—Cloud provider provides software facilities which are not available in single system. For eg :Microsoft Office 365
- Cloud Platform as a Service (PaaS)—this service model provide various platform, which cannot be readily available due to limited computational power of the standalone system. for e.g.: Google App Engine
- Cloud Infrastructure as a Service (IaaS)—Resource computing, Data storage, Network Traffic Controlling, Memorization. for e.g.: Google web services

The deployment models as shown in Fig 1.1 are as follows:

- Private cloud—owned or leased by an enterprise
- Public cloud—mega-scale infrastructure, sold on demand
- Hybrid cloud—Composition of two or more private and public clouds
- Community cloud—Shared infrastructure for specific organisation



**Fig 1.1 Cloud Computing Deployment Models**

The system reliability is the measure of number of potential points of failure and the redundancy (or lack thereof) around those points. Cloud computing is very efficient in shifting the reliability of a system from hardware functionality, to relying on the availability of services (SaaS, PaaS, IaaS). The virtualization feature of cloud at different levels like storage, server, network has also improved the reliability and throughput of the system. However, with the use of cloud computing many cloud handling and maintenance challenges have come into existence. One of the major challenges among this is security over cloud. There are sudden and fast changes in the customer services, security requirement in cloud. There is recent interest in the computing service, known as container based computing, that speeds up the computing process and also bridge the gap of security requirement. Thus, these rapidly changing security threats motivated to find new methods to ensure security over cloud environment.

## 1.2 CLOUD SECURITY ARCHITECTURE

The rapid evolution of cloud computing services, has promoted many organization to migrate towards consuming the cloud services, especially towards public cloud services, which is under the control of third party Cloud Services Provider (CSP). Due to which, many security issues gets associated with these services faced by both the cloud service providers and the customers. Thus, the responsibility of resolving these issues go both ways i.e. provider must ensure a secure infrastructure (the CSP must ensure the data isolation and data segregation) and customer must ensure security of their data through various authentication scheme and cryptographic algorithm.

The model for an ideal and secure Cloud environment [1] consists of following features:

### 1. Authentication:

Authentication is the process of reconciliation of evidence for confirming the user's identity. It aims at identifying the user and validating user's identity who they claim to be through evidence they provide like password. One-time passwords, x.509 certificates, and device fingerprinting are several strong user authentication methods. Federated identity permits a user to work in field, known to be Software-as-as-Service (SaaS) application, and getting authenticated in another domain, such as a corporate Identity Management (IdM) system.

### 2. Authorization

Authorization is the second step which determines the access privileges given to the user, it determines which cloud resources user can access once the user identity is established and authenticated successfully. Nowadays, emphasis is being laid on centralizing the authorization policy decisions irrespective of user's physical location.

### 3. Account Management

Accountability is the ability to inspect the user logging activity accessing behaviour of an individual user within a cloud system and identification of the particular user. It's mandatory for each user to have an account commonly for audit purposes and licensing. Account management refers to protocols and standards such as SCIM that maintains the synchronization of user cloud account with existing enterprise systems.

#### **4. Audit Logging**

To manage the operational assurance, enterprises use two basic methods: system audits and monitoring. There are two ways, which is being exercised by the cloud customer, the cloud vendor, or may be both. It depends upon cloud deployment and asset architecture.

- A system audit is a periodic event aims to evaluate security.
- Monitoring is a continuous ongoing activity that inspects either the users or the system, for various attacks

The ability of an organisation to keep the track of users logging and surfing activities, is a major concern from security perspective. Auditing records and maintaining the logs of user activities serves important evidence in cyber forensic activities of cloud. Several breaks down in authentication sessions or unauthorized services access will spotlight potential threats, terror and fraud related activities. In addition, enterprise requires audit trails for the successful verification and authorization of the enrolled users, and ensures that only legitimate clients have privilege access to the authorize systems.

### **1.3 NEED FOR AUTHENTICATION IN CLOUD**

Cloud computing infrastructure security is highly influenced by the cloud selection for deployment i.e. to use a private cloud or a public cloud. The security infrastructure of Private cloud almost exactly duplicates traditional IT security architecture and shows similarity to a private extranet implementation. However, the public cloud infrastructure requires the organization attention towards security architecture and its synchronisation with the CSP's network. In every case, implementation of a secure cloud effectively reduces the risk to confidentiality, integrity, and availability; ensure proper access control (authentication, authorization, and auditing.) and protected data storage. [3]

Frequency of security breaches on (i.e. Twitter, LinkedIn) has forced the IT departments to pay closer attention to authentication. Ubiquity of mobile devices has resulted a tremendous increase in the usage of online apps and also increasing the number of options for assisting user in authentication. In the past few years, the popularity of phone-as-a-token solutions has overtaken

One-time password hardware tokens in terms of new and refreshed deployments.

The authentication method has “enhanced and morphed from traditional tokens to USB devices to smart cards to biometric traits reader, soft tokens and one time password.” Moreover, study and analysis of biometric traits for authentication is widely accepted and considered as high level of assurance, including form factors like typing rhythm, voice recognition, face topography and iris structure.

Migration towards the authentication services delivered by cloud is emerging as a popular and widely adopted service and having the maximum traction among small and mid-sized businesses and industries where TCO is a more significant consideration. [2] It has been predicted that by 2018, more than 70% of enterprises will adopt cloud-based services – up from less than 20% today.

For hybrid cloud the authentication depends upon how the cloud is implemented i.e. if it is implemented using private cloud via a VPN then authentication is similar to other intranet networks, but authentication becomes complicated with pure public cloud. Public cloud is generally under the control of third party and remains external to the enterprise that uses it.

Authentication over public cloud is major security concern, the flow of authentication and credentials validation while logging on to a web application should be audited and account management is done. For example, Salesforce allows you to sign a federation agreement with them thereby who you are a given permission to register and access with their private login id and password. Unfortunately, identity federation is structured for simple Web application use cases only. Therefore, automatic backup of your local disk to a cloud service, which is public. It can't be done easily using the corporate credentials for your authentication to the service provided. You are not even able to permit to access local application that is SQL database. This is very apparently put thrust on federated corporate identity to secure these issues. These all security consequences are raised by cloud environment.

## 1.4 RELATED WORK

So far, significant research work has been carried out to provide adequate security to the cloud environment. These existing security mechanisms still lacking behind to provide adequate security to clouds. T S Khatri and G B Jethava et al [4] conducted a survey on various cloud computing security paradigms and analyses the new issue faced by the cloud environment especially the public cloud.

Amlan Jyoti Choudhury, Mangal Sain, Pardeep Kumar [41] proposed a authentication framework, which is strong enough to ensure mutual authentication, identity management and session key establishment in cloud environment. The identity management is ensure by managing the ID-table for the registered user and performs mutual authentication of the user using OTP scheme and smart card services. This approach helps in reducing the cloud security concerns to a tremendous level but it fails in case the security of the backend server if compromised. However, this approach shows a sting resistant toward many security attacks such as phishing, guessing and replay.

G.SudhaSadasivam, K.AnithaKumari et. al [42] proposed a technique to enhance security characteristic by using triangle properties. It proposed a two server model. In which the registered login details of the user is analysed and quarantined into many unit and stored at backend server and the authentication server. Authentication server accumulates the information from every backend servers and performs authentication.

An agent based user authentication scheme is proposed by Faraz Fatemi Moghaddam, Touraj Khodadadi, [47] in which process of authentication is carried out using agent. The first agent is responsible for authenticating the secure channel, second agent performs client authentication and third agent manage and stores the key used in above authentication process.

Many Id-based user authentication schemes were also proposed [43-46]. In 2004, dynamic ID-based remote user authentication scheme was proposed by Das. Their Scheme enables users to select and change their passwords anytime without maintain any verifier table. However, Wang et al. [44] pointed out the flaw in Das et al.'s scheme and stated that it has no resistant toward main in middle attack and user can false fully legitimate the user password. Unfortunately, Lee

et al. [41] found that Wang et al.'s scheme still had some security issues. That is, Wang et al.'s scheme cannot prevent from the message alteration attack and impersonation attack. Besides, we find that Wang et al.'s scheme has high computation cost and time complexity. Thus, it is not suitable for cloud computing.

Though many significant and strong authentication schemes has been proposed by many researchers and analysed, the rapid growing of cloud services has introduced many new security issues like data isolation, data segregation, web authentication and legitimating the user's identity successfully as in intranet environment with low computational cost and setup cost.

Thus, our proposed work aims at resolving certain cloud security issues such as authentication, data isolation and maintaining data integrity.



# DNA CRYPTOGRAPHY AND SEQUENCE ALIGNMENT

---

## 2.1 INTRODUCTION

The science of DNA sequencing deals with identifying the order of nucleotides, which remains same as per DNA molecule. It comprises of several methods or technology that aims to determine the order of the four bases— cytosine, adenine, guanine, and thymine—in a strand of DNA. In the early 1970s, academic researchers obtained the first DNA sequence based on two-D chromatography. The sudden evolution of DNA sequencing and Sequence Alignments methods has greatly induced the field of biological research. There is many application area of DNA sequencing such as forensic biology, virology, biotechnology and biological system. The ability of DNA to carry massive parallel information has simulated the use of DNA to hide the information and utilizing then for cryptographic purpose. Thus, giving rise to new domain ‘DNA cryptography’ and which is continuously growing with the exploring of DNA computing.

## 2.2 DNA CRYPTOGRAPHY

There is lot of research going on since mid 90’s, in the field of DNA Computing. Today, the science of DNA computing posses a high level computational ability and have the potentiality to solve huge and complex mathematical problems. The massive parallel nature of DNA and ability to bear the extraordinary information density is a key feature which has been efficiently utilized by the researchers for data hiding purposes and all sort of cryptographic purposes, giving rise to new field known as DNA Cryptography. There is very fast evolution in DNA cryptography and it continuously growing with the advent of fields such as DNA computing.

### 2.2.1 Biological Background of DNA

DNA is acronym for deoxyribonucleic acid which is germ plasma of all lifestyle. It is a biological macromolecule and is made up of nucleotide. There are Nucleic acids, which consists of a chemical string of interlinked attribute called nucleotides. Each nucleotide comprises of three things: a sugar (ribose in RNA, deoxyribose in DNA) a phosphate group and which act as

integral unit of nucleic acid strand, and attached to the sugar is one of a set of nucleobases. These nucleobases are accountable for double helical structure of DNA which participates in base pairing of DNA strands to form higher-level structure as secondary and tertiary [5][6].

The four nucleotide bases of DNA strand is shown in the table 1 representing the four nucleotide bases of a DNA strand, which is covalently linked to a phosphodiester backbone.

A	Adenine
T	Thymine
G	Guanine
C	Cytosine

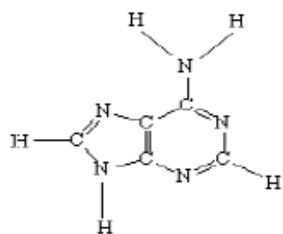
**Table 1 DNA Nucleotide Base**

From figure 2.1(as per their chemical structure), the four nucleotides A, C, G, T can be divided into two classes:

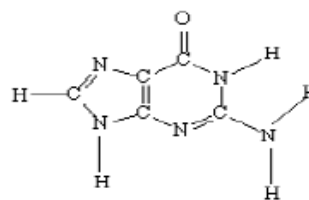
1. Purine R= {A, G} and Pyrimidine Y= {C, T}
2. Amino Group N = {A, C} and Keto Group K= {G, T}.

Apart from these divisions, further bifurcation can be made on the basis of the hydrogen bonds, i.e. how strong the bond is? Strong H-bonds S= {G,C} and Weak H-bonds W={A,T}.[8]

## PURINES

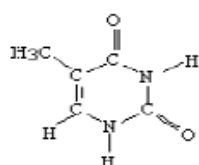


ADENINE

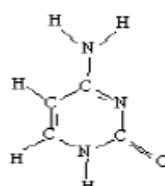


GUANINE

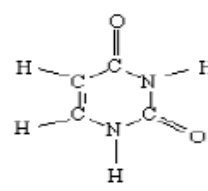
## PYRIMIDINES



THYMINE



CYTOSINE



URACIL

**Fig 2.1 Structure Of Purines and Pyrimidnes**

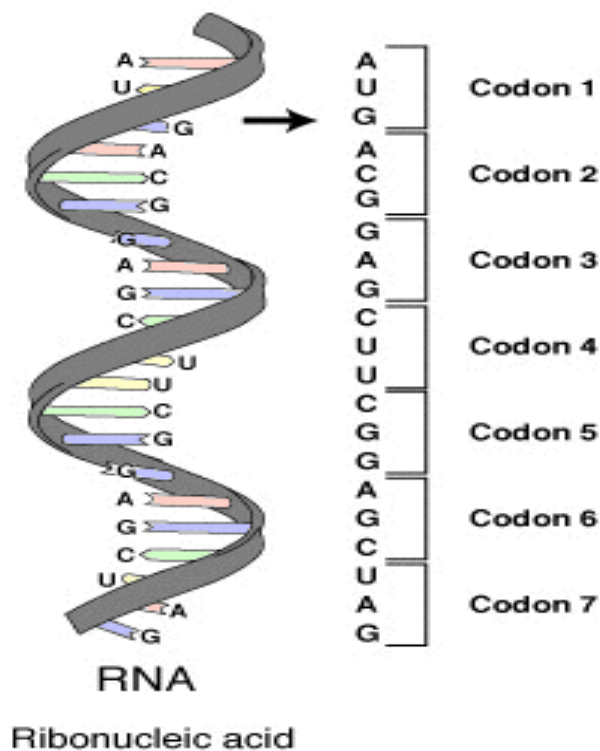
A gene is DNA sequence which carries the genetic information. Within a gene, a messenger RNA sequence can be defined on the basis arrangement of bases in a DNA strand. The translation and transcription technique are collectively known as genetic coding. [9][10] It defines the relationship between the amino-acid sequences of proteins and the nucleotide sequences of genes. The word “Genetic Code” comprises of ‘words’ of three-letter called codons composed from a sequence of four nucleotides bases (e.g. ACT, CAG, TTT). In transcription, a DNA segment is transformed into messenger RNA (mRNA) by RNA polymerase, which exits the nucleus and enters into the body of a cell. In translation, the encoded information in mRNA is decoded by ribosome and assembles amino acid into protein chains.

### 2.2.2 DNA Sequencing

DNA sequencing is the process of determining the precise order of nucleotides within a DNA molecule. Any method or technology that defines the ordering of four nucleotide bases i.e. adenine, guanine, cytosine, and thymine in a DNA strand. Sequencing methods of DNA has

greatly fastens the medical and biological research. The field of natural sequence pattern along with chemical classification and complementary genetic coding is used to protect or hide the message.

Two sequences can be said as complementary sequence, if their base position is complementary to each other and also when the order is reverse. The arrangement of series of codon in a mRNA molecule is shown in Fig 2.2 describing the complementary properties of nucleotide base. For example, the complementary sequence of ATGC is TACG. Thus, if one is a sense strand then the other is antisense strand and shows complementary behaviour to the other strands.



**Fig. 2.2 Series of codon in mRNA molecule.**

There is one special property for DNA sequences i.e. the original DNA sequence and the faked DNA sequence will almost look like the same. And, there are also a large number of DNA

databases which are publicly available. By using these facts, in this paper a new methodology is formed for encrypting messages using DNA sequences.

DNA sequences offer a unique method of encrypting messages or information. The main advantage of DNA sequences is they are composed of letters which are meaningless for most people. The DNA sequence is a combination of A, C, G and T base pairs. [31]

Using these properties of DNA sequences, three complementary rules can be formed and subsequently be used to generate fake DNA sequences. DNA sequence serves as an ultra-compact information storage medium which stores a large amount of data in compressed form. A single gram of DNA contains 1021 DNA bases = 108 tetra bytes.

Thus, these characteristic of DNA:

- Massive parallel computing
- Large data storage
- Information carrier (mRNA)
- Genetic coding
- Generation of faked and random DNA sequences
- Searching complexity of a particular DNA sequence in large database

Has increased the possibility of using DNA and gives a prominent direction in cryptographic research.

There are several DNA-based algorithms that have been practically applied for cryptography purposes. Kang Ning proposes a method in which sender uses the original DNA sequence to encode its secret message and performs transcription and translation obtaining a protein which act as a public key for the receiver. Ning also proves that this cryptography method is secure against many intruder attacks like replay attack, brute force attack though he acknowledges that the encryption complexity increase with key size. [10]

Debnath Bhattacharyya proposed a new data hiding methods based upon DNA complementary rules and message indexing. In this indexing of a random DNA sequence is done which is used as a reference for encoding the message in DNA sequence and during decryption process one requires the DNA string and Index mapping to obtain the original message. [11]

In 2010, H.J. Shiu, K.L. N proposed a more robust method for hiding data.[32] He introduced the insertion method, substitution method and complementary pair method for hiding data in DNA sequence. In the insertion method both the reference DNA sequence and the secret message are assembled from the sequence and the secret message after decomposing. In the Complementary Pair Method the complementary rules are used to encode the secret message. In the substitution method, another letter is substituted for an existing letter decided by the algorithm substitution rule.

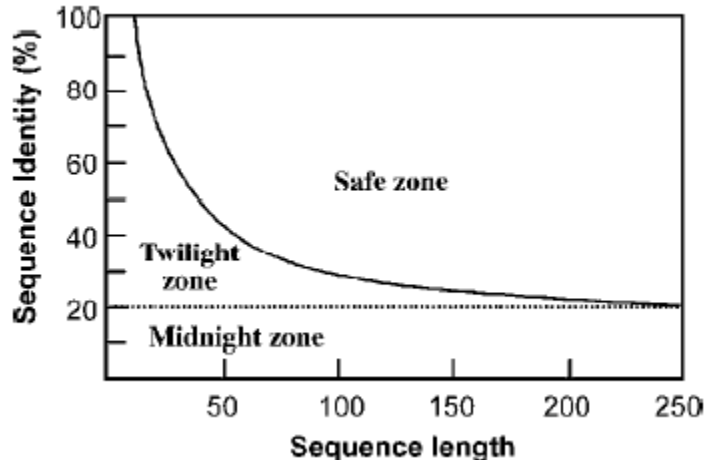
The DNA-crypt algorithm has also been used for image cryptography. Qiang Zhang, Ling Guo proposed a new scheme of encrypting image using DNA sequences.[37] The proposed approach defines two new mathematical operation on DNA sequences i.e. addition and subtraction operation which helps in combining the encoded matrix block of DNA sequences and then using complementary rules for the output of added matrix block by using chaotic dynamical system(Logistic Map).

Jin-Shiuh Taur et al. [38] proposed a method for improving the effectiveness of the substitution method, known as Table Lookup Substitution Method (TLSM), this methods enhance the message hiding capacity twice. In TLSM, the extended the complementary rule definition and introduce a 2-bit rule table instead of 1-bit rule table. Thus, while encoding allowing two bits of secret message to be encoded.

## **2.3 SEQUENCE ALIGNMENT**

In bioinformatics, sequence alignment is one of the integral techniques. It implies the arrangement of the DNA, RNA, or protein in order to judge the similarity among the sequences based upon functional, structural or evolutionary relationship. The evolutionary analogy of a nucleotide or amino acids sequence is given by the degree of sequence conservation, whereas the degree of variation shows the deviation that have occurred in the form of insertions, deletion and substitution during evolution. We need a minimum of two sequences to be aligned known as pairwise alignment and maximum varies to thousand. The Alignment of sequences of nucleotide and amino acid are illustrated in row major matrix form.

Sequence homology is an important phenomenon in sequence alignment analysis. Two sequences are said to be homogenously related when they are descended from a same evolutionary origin. A related but different term is sequence similarity, which is the percentage of aligned residues of a nucleotide or amino acid that are alike in physiochemical properties such as charge, size and hydrophobicity. Sequence homology is an illation obtained from sequence similarity comparative analysis about common ancestral relationships. Whereas, similarity is a direct result of experimental observation from the sequence alignment. Sequence similarity is quantitative measure and homology is qualitative inference. For example, if two nucleotide sequences share 45% similarity, then we cannot say they share 45% homology. They are either homologous or nonhomologous. Inferring homologous relationships from a particular similarity level depends upon type of sequence and sequence length. Length of nucleotide or amino acid sequence is also an important factor. Longer sequences needs lesser cut-offs for inferring homologous than shorter sequences.



**Fig 2.3 Three zones of nucleotide sequence alignments.**

The zones of nucleotide sequence alignment is shown in the Fig 2.3, if the percentage sequence identity falls in safe zone, then the nucleotide sequences are homologous. If the percentage sequence identity falls in twilight zone, then the homologous relationships are less certain and in the midnight zone, it cannot be determined reliably.

### 2.3.1 Sequence Similarity versus Sequence Identity

The percentage of sequence identity or sequence similarity is given by the number of identical or similar amino acids residues respectively, when compared to the total number of amino acids in the protein. There is a variation between identity and similarity. Identity is the degree of correlation between 2 un-gapped sequences, which implies an exact match at a particular region of the amino acids or nucleotides. The degree of correlation among the two sequences when they are compared is known as similarity. This means that few properties such as charge or hydrophobicity of amino acids or nucleotides residue are related but not exactly same. The degree of similarity or identity is mediated by the distance for two sequences. Generally, a +1 is assigned for a match, 0 is assigned for a mismatch and -1 for a gap (indel), and the distance score is given by summing all these alignment scores.

Sequence similarity or sequence identity is calculated in two ways:

1. In the first method, the normalization is done using overall sequence length of both the sequences. The percentage similarity score and identity score is calculated using eq1 and eq2 respectively

$$\text{Score}_s(\%) = [(\text{Length}_s \times 2)/(\text{Length}_a + \text{Length}_b)] \times 100 \quad \text{eq(1)}$$

$$\text{Score}_i(\%) = [(\text{Length}_i \times 2)/(\text{Length}_a + \text{Length}_b)] \times 100 \quad \text{eq(2)}$$

Where, **Score<sub>s</sub>** is the percentage sequence similarity

**Score<sub>i</sub>** is the percentage sequence identity

**Length<sub>s</sub>** is the number of aligned residues with similar characteristics

**Length<sub>a</sub>** and **Length<sub>b</sub>** are the total lengths of each individual sequence.

2. In the second method, the percentage of identical or similar residues is calculated by the full length of the smaller sequence and is given by eq 3

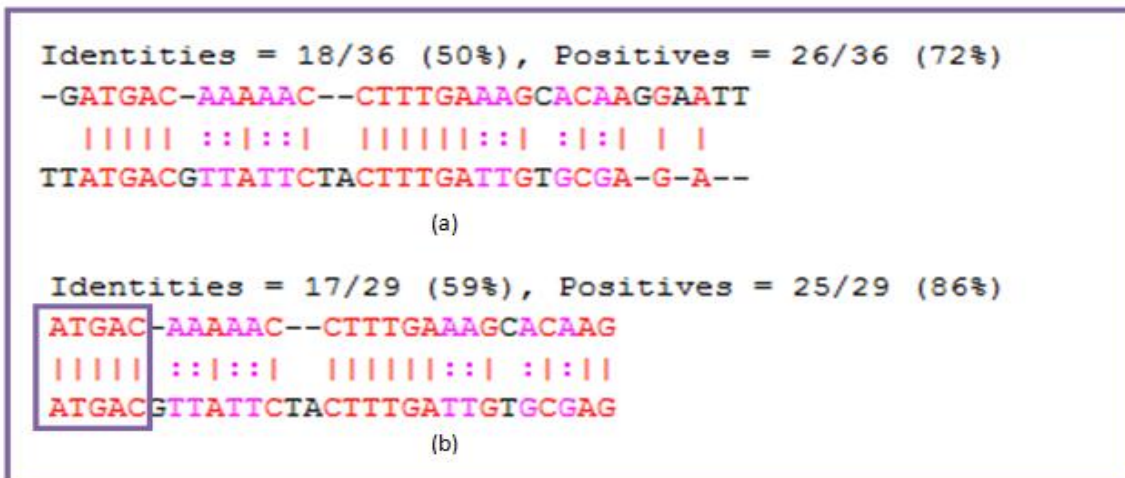
$$\text{Score}_{I(S)}\% = \text{Length}_{i(s)} / \text{Length}_a\% \quad \text{eq (3)}$$

Where, **Length<sub>a</sub>** is the sequence shorter in length



The main aim of the pairwise sequence alignment is to determine the optimized pairing of nucleotide sequences, and to attain a maximum correlation score among the sequences. To attain this agenda, one nucleotide sequence needs to be aligned relative to the other (reference sequence) in order to discover the position where maximum correlation is seen. There are two types of sequence alignment strategies:

**Global Alignment:** In global alignment, the two nucleotide or amino acids sequences to be aligned are considered to be generally similar over their entire length as shown in Fig 2.4(a). Alignment is carried out from initial to end of both the sequences in order to determine the most optimized alignment across the entire length between the two nucleotide sequences. This technique is more useful for aligning two nearly related sequences of approximately the same length. The sequences having variable length and diverge, this method may not give the best possible results as it fails to identify local regions having high similarity.



**Fig 2.4 (a) The global alignment of nucleotide sequence (b) The local alignment of nucleotide sequence**

**Local Alignment:** In local alignment, the two nucleotide or amino acids sequences to be aligned does not necessarily be similar over the full length of the sequences. It discovers the local regions which posses the maximum level of similarity between the two nucleotide sequences and alignment of these regions without considering the alignment probability of the rest of the sequence regions as shown in Fig 2.4(b). This approach of local alignment can be utilized for obtaining more divergent sequences with the aim of discovering the all possible conserved

patterns in DNA or protein sequences. The length of aligned sequences can be different in case of local alignment. This methodology is more appropriate for the alignment of divergent nucleotide sequences that are similar, also known as domains or motifs.

The Alignment algorithms, both global and local are similar. They differ in the optimization strategy adopted for aligning the sequence residue. There are three methods for alignment:

- i. Dot matrix method
- ii. Dynamic programming method
- iii. Word method

## **2.4 DYNAMIC PROGRAMMING METHOD**

In Dynamic programming method the optimal alignment is done by matching the two nucleotide sequences for all possible pairs of characters between the two sequences. It is conceptually alike to the dot matrix method in which we create a two dimensional alignment grid. It helps in obtaining the alignment in a quantitative way by transforming the dot matrix into score matrix to calculate the matches and mismatches between the aligned sequences. The best and optimized alignment is obtained by searching for the set of highest scores in the score matrix.

Dynamic programming calculate the score and the optimized path by constructing a two-dimensional matrix whose axes X and Y represents the two nucleotide sequences to be matched. For calculating the score one row is considered at a time. Starting with the first row of one sequence, through which the entire length of the other sequence is examined, followed by glancing over the second row and the matching scores is calculated. The scanning of the second row considers the first row, calculated score as an initial score and the best score is placed into the bottom right corner of an intermediate matrix. Until, all the cells of the score matrix is filled this process is iterated and finally the score accumulation is done along the diagonal of the score matrix from upper left corner to lower right corner. Now, the path is discovered for the accumulated score which represents the optimal alignment. This is done by backtracking the matrix in reverse direction from lower right corner to the matrix origin in upper left corner. The path showing the maximum total score is considered as best path. The introduction of insertion and deletion gap between the nucleotide sequences results in the change in direction of path towards horizontal or vertical at a certain points.

**Gap Penalties:** While performing the optimal alignment between the two nucleotide sequences often involves applying gaps which shows insertions and deletions. Introduction of insertions and deletions gaps in natural evolutionary processes has increased the computational difficulty. However, deciding the penalty values can be more or less random because there is no predefined evolutionary theory to calculate the optimized cost for introducing insertions and deletions. Setting low penalty values, results in matching of even non-correlated sequences with high similarity score, because of introduction of numerous gaps between the sequences. Whereas, keeping the penalty value high, appearance of gaps is rare and thus fair alignment cannot be attained, which is inaccurate. Therefore, a set of penalty values have been fixed that appropriately suits every pairwise alignment purpose.

**2.4.1 Dynamic Programming Method for Global Alignment:** The pairwise global alignment is done using Needleman-Wunsch algorithm using dynamic programming. The alignment is carried out over the entire length of the two nucleotide sequences to obtain the highest total score. The winning path for global alignment goes from bottom right corner to top left corner of the scoring matrix. The only disadvantage of global alignment is concentrating on achieving a highest score for the entire-length of nucleotide sequence alignment results in the risk of losing the best local similarity. This strategy appropriate for the alignment of those two nucleotide sequence almost same in length. This algorithm does not produce optimal alignment for sequences having different domain structure or if it is a diverging sequence. Few web servers dedicated to global pairwise alignment is GAP.

## Needleman- Wunsch Algorithm

Input: Two nucleotide sequence  $A = \{a_1, a_2, a_3, \dots, a_m\}$  and  $B = \{b_1, b_2, b_3, \dots, b_n\}$

The scoring scheme consists of substitution score {Match [+1], Mismatch [-1]} and Gap[0]

Step 1 Initialization of Scoring matrix S

$$T_{i0} = T_{0j} = 0 \quad \text{for } \{0 < i < m \text{ and } 0 < j < n\}$$

Step 2 Calculation of score T [i, j] for each cell of the scoring matrix

$$T_{0j} = d * j$$

$$T_{i0} = d * i$$

$$F_{ij} = \max (F_{i-1,j-1} + S (A_i, B_j), F_{i,j-1} + d, F_{i-1,j} + d)$$

Step 3 On finding highest score in the matrix S, back trace the path using back pointers to obtain the optimal alignment.

**2.4.2 Dynamic Programming Method for Local Alignment:** In global alignment, the divergence level between the two nucleotide sequences cannot be easily determined. The sequence lengths of the two sequences may also be unequal. Smith-Waterman techniques evaluate the relatedness between two sequences locally using dynamic programming method. It only covers the part of DNA sequence. It aims at comparing the segments of all possible lengths and optimizes the similarity measures between two strings of nucleotide and protein sequence. The scoring scheme consists of match, mismatch, substitution, insertion/deletion gap penalties. In this algorithm, for a matching residue a positive scores is assigned and zeros for a mismatching residue, no negative scores is assigned. In this the back tracing path may begin and end intermediately along the main diagonal. It begins with the position having maximum score and back trace it diagonally up to the left till a cell having zero value is reached, insertion of gap is done if necessary. This algorithm introduces new gap penalties i.e. gap initiation (cost for each continuous

run of gap) and gap extension (cost for each piece of gap). The aim of local alignment is to obtain the highest alignment score locally, which may be at the expense of the highest possible overall score for the entire-length of sequence alignment. This strategy is appropriate for the alignment of those two nucleotide sequence which are divergent and have different domain structure. Few web servers dedicated to local pairwise alignment which includes SIM, SEARCH and LALIGN.

### Smith-Waterman Algorithm for Local Alignment

Input: Two nucleotide sequence  $A = \{a_1, a_2, a_3, \dots, a_m\}$  and  $B = \{b_1, b_2, b_3, \dots, b_n\}$

The scoring scheme consists of substitution score {Match [+1], Mismatch [-1]} and Gap [0]

Step1 Initialization of Scoring matrix S

$S_{i0} = S_{0j} = 0$  for  $\{0 < i < m \text{ and } 0 < j < n\}$

Step 2 Calculation of score S [i, j] for each cell of the scoring matrix

$$S(i, j) = \max [S(i-1, j-1) + \sigma(S_1(i), S_2(j))]$$

$$\max [S(i-1, j) + \text{gap penalty}]$$

$$\max [S(i, j-1) + \text{gap penalty}]$$

Step 3 On finding highest score in the matrix S, back trace the path using back pointers to obtain the optimal alignment.

# CHAOS AND CRYPTOGRAPHY

---

### 3.1 INTRODUCTION

In past few decades, chaotic system broadly used in cryptography system. Chaos a Greek word, which implies uncertainty and is under the non-linear dynamic system. Chaotic systems are characterized by its high sensitivity which is dominated by its initial conditions, continuous broad-band power spectrum and its random behaviour. Chaotic systems are very integral part of cryptographic ingredients, such as encryption, decryption of message and also plays import role in modulation and compression. The hypothesis for self synchronization of chaotic oscillations [13] has resulted in wide application of chaos in cryptography.

Chaos theory is used in several disciplines like meteorology, physics, biology, economics and engineering. Chaotic systems like Henon map, Tent map, Rossler attractor, Logistic map and Piecewise linear chaotic map are known for their randomness and unpredictability. These systems exhibits randomness property dependent upon the initial parameters which can be seen in the orbit of these map in the attractor field. Chaotic systems are prone to attack as they transiently predictable but, indeed not. The period of time during which the behaviour of a chaotic system can be determined efficiently depends on three parameters: the amount of uncertainty a system can tolerate in the forecast, the accuracy of the system current state and a time scale depending on the dynamics of the system, called the Lyapunov time[14][15]. Some examples of Lyapunov times are: chaotic electrical circuits, about 1 millisecond; weather systems, a few days (unproven); the solar system, 50 million Werndl, Charlotte (2009).Chaotic system based on confusion and diffusion was developed in 1989, which has been used in cryptographic system but the authenticity of the chaotic map used in that was not proved[16] The initialization parameter used in chaotic maps should be the real numbers [17],which can be used as keys in cryptographic algorithms. A chaotic systems characteristic like sensitivity, random-look nature, uncertainty, reconstruction after filling in the multimedia resulted in its popularity and is widely used in cryptographic algorithm such as block ciphers, hash functions and pseudorandom number generator [21].

## 3.2 CHAOS THEORY

Generally "chaos" means "a state of disorder"[19] or randomness but, a dynamical system is considered can act as chaotic system, then three properties must be establish [20].

- It must be topologically mixing.
- It must be sensitive to initial conditions,
- It must have dense periodic orbits

Chaos theory has gain popularity in the field of computer science and has potential application in cryptography. Chaos theory becomes more effective, there is an amalgamation between DNA computing and chaos theory. This leads to a more prominent method in the area of encryption and decryption and other multimedia techniques such as audio, video and other lightweight and heavyweight messages. [18]

A chaotic system exhibits a simple, non-linear and dynamical behaviour. Its deterministic behaviour is determined by how the current state is uniquely mixed and by the density and randomness of the periodic orbit. The most prominent behaviour of chaotic system is having a great sensitivity for its initialization parameter i.e. a slight variation in initial values might generate large differences in the results. [22]

In contrast to stochastic system, chaotic system behaves differently compared to stochastic but shows similar behaviour as deterministic system. Error in deterministic either grows exponentially or remains small. Typically for measuring the deterministic behaviour one selects an embedding dimension and starts investigating in the direction of error propagation between two nearby states. If increasing the dimension helps in obtaining a deterministically identifying the error, then the analysis of the deterministic system is successfully done. However, the computation complexity increases because of increase in computation time and amount of data required. There is challenge to differentiate between chaotic systems, deterministic and stochastic systems. They have also been discussed and assumed that they might be observationally equivalent. [23]

Since there is always a mathematical expression to determine the system evolution and from the above definition of the deterministic and dynamical system, we can infer that randomness is

allowed and a slight variations in the parameters, results in number of solutions because of deviation in dynamic differential equations. The chaotic behaviour can be observed in many dynamic systems like electronic systems, fluid dynamics, weather, climate and lasers. [24]

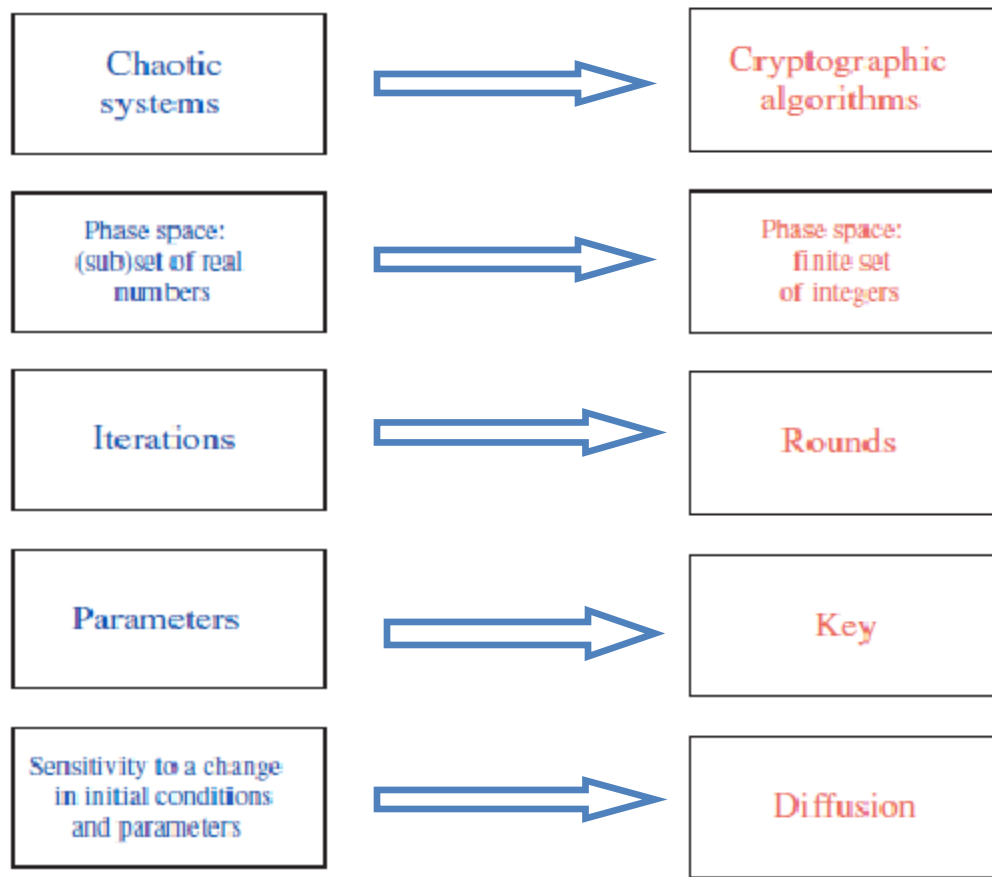
Generally, chaotic maps use large values of real numbers. The most prominent characteristics of chaotic systems are:

1. **Randomness but deterministic in nature:** Chaotic maps show purely deterministic nature but are random in their behaviour. Hence, for the same initial parameters the system generates the same output values repeatedly. Moreover, chaotic systems can be determined by differential equations or iterative mappings.[22]
2. **Sensitivity:** The initialization parameters results in ignition of the systems. The behaviour of dynamical system changes entirely with a slight variation in the initial parameter. [25].Thus, changing the initial variable for e.g. 0.01 to 0.00001, will not produce the same key stream obtained from the previous input initial value in chaotic systems.
3. **Uncertainty:** In chaotic systems, even if the current state is known it is difficult to determine the next state of the system. Thus, prediction of future state is quite hard in chaotic systems.

### 3.3 CHAOS-BASED CRYPTOGRAPHY

There is contrasting relationship between Chaotic system and cryptographic and their is urgent need to understand it. It is shown in fig. 3.1.The conventional cryptographic systems uses integer number system for key stream generation whereas, chaotic based system uses floating point number systems for key stream generation.[26] The three common primitives of cryptography where chaotic systems have popular applications are block cipher, hash functions, and pseudorandom number generator. [25]





**Fig 3.1 Relationship between chaotic systems and cryptographic algorithms**

**Block Cipher Based on Chaotic Systems:** In block cipher technique, a transformation function is used which maps the plaintext bits to cipher text bits of same size using the secret key. Chaotic cryptosystems have high computational speed with minimum cost, and increases its application in many traditional and block ciphers for multimedia data encryption. [27]

**Hash function Based on Chaotic Systems:** Secure Hash Algorithm (SHA) is the most widely used hash functions algorithm in cryptography. Many security protocols like PGP uses hash functions for the integrity of data. Chaotic hash function is widely used approach in cryptography algorithm aims at designing a new hash function which is more collision resistant compared to SHA-1. [28]

**Random Number Generators Based On Chaotic Maps:** The results obtained from chaotic systems are often unpredictable and uncertain and this property of chaotic systems is used to design pseudo random number generators. The output of PRNG is used as a key stream in many cryptographic encryption algorithms. [29][30]

### 3.4 CHAOTIC MAPS

A chaotic map is a graphical representation of an evolution function that exhibits some sort of chaotic behaviour. The parameters of the chaotic maps are either discrete-time or a continuous-time parameter. Discrete maps are usually taken in the form of iterated functions.

#### 3.4.1 One-Dimensional Chaotic Maps:

A 1-D chaotic map deals with only single physical quantity. Traditionally, the input or order values are across horizontal axis and the corresponding output value on the vertical axis. List of some 1-D chaotic maps are:

- **Tent Map:** the tent map with parameter  $\mu$  is the real-valued function  $f_\mu$  defined by

$$f_\mu = \mu \min(x, 1-x) \quad \text{eq(4)}$$

- **Gauss Map:** the Gauss map or mouse map, is a nonlinear iterated map of the reals into a real interval given by the Gaussian function:

$$X_{n+1} = \exp(-\alpha X_n^2) + \beta Y_n \quad \text{eq(5)}$$

- **Logistic Map:** The logistic map is a quadratic polynomial which shows how complex, chaotic behaviour can arise from very simple non-linear dynamical equations given by

$$X_{n+1} = rX_n(1-X_n) \quad \text{eq (6)}$$

#### 3.4.2 Two-Dimensional Chaotic Maps:

A 2-D maps deal with more than one physical quantity. 2-D chaotic map exhibits the evolution function in a three dimensional space where x and y axis indicates the equation of chaotic maps and z-axis is a temporal axis. List of some 2-D chaotic maps are:

- **Duffing Map:** The Duffing map is a discrete-time dynamical system. It is an example of a dynamical system that exhibits chaotic behavior. The Duffing map takes a point  $(x_n, y_n)$  in the plane and maps it to a new point given by eq(8). The map depends on the two constants  $a$  and  $b$ . These are usually set to  $a = 2.75$  and  $b = 0.2$  to produce chaotic behaviour. It is a discrete version of the Duffing equation.

$$\begin{aligned} X_{n+1} &= Y_n \\ Y_{n+1} &= -bX_n + aY_n - Y_n^3 \end{aligned} \quad \text{eq(8)}$$

- **Gingerbreadman map:** It is a chaotic 2D map in the dynamical system given by

$$\begin{aligned} X_{n+1} &= 1 - Y_n + |X_n| \\ Y_{n+1} &= X_n \end{aligned} \quad \text{eq(9)}$$

- **Hénon map:** It is a discrete-time dynamical system, that exhibit chaotic behaviour. The Hénon map takes a point  $(x_n, y_n)$  in the plane and maps it to a new point using

$$\begin{aligned} X_{n+1} &= 1 + Y_n - aX_n^2 \\ Y_{n+1} &= bX_n, \text{ where } n=0, 1, 2, \dots \end{aligned} \quad \text{eq(10)}$$

- **Standard map:** It is also known as the Chirikov standard map. It is an area-preserving chaotic map from a square with side  $2\pi$  onto itself. It is constructed by a Poincaré's surface of section of the kicked rotator, and is defined by:

$$\begin{aligned} p_{n+1} &= p_n + K \sin(\theta_n) \\ \theta_{n+1} &= \theta_n + p_{n+1} \end{aligned} \quad \text{eq(11)}$$

### 3.5 HENON MAP

The Henon map is a 2-D iterated map which shows the chaotic behaviour proposed by a French astronomer Michel Henon in 1976 as simplified model of the Poincare map for the Lorenz model. The mathematical equation eq 12 for the Henon Chaotic map helps in generation of pseudo random binary sequence.

$$\begin{aligned} X_{n+1} &= 1 + Y_n - aX_n^2 \\ Y_{n+1} &= bX_n, \text{ where } n=0, 1, 2, \dots \end{aligned} \quad \text{eq(12)}$$

Where ‘a’ and ‘b’ are two bifurcation parameter which dominates the behaviour of Henon Map. ‘The measure of the rate of area contraction is determined by ‘b’. For b=0, the Henon map reduces to a quadratic equation and its graphical representation becomes analogous to logistic map. Henon map is defined for the real values of  $X_n$  and  $Y_n$  and the contraction property of Henon map is independent of these co-ordinates. Henon map is defined for discrete time domain. The dynamic behaviour of the system is determined by ‘a’ and ‘b’ and these parameters are of prime importance.

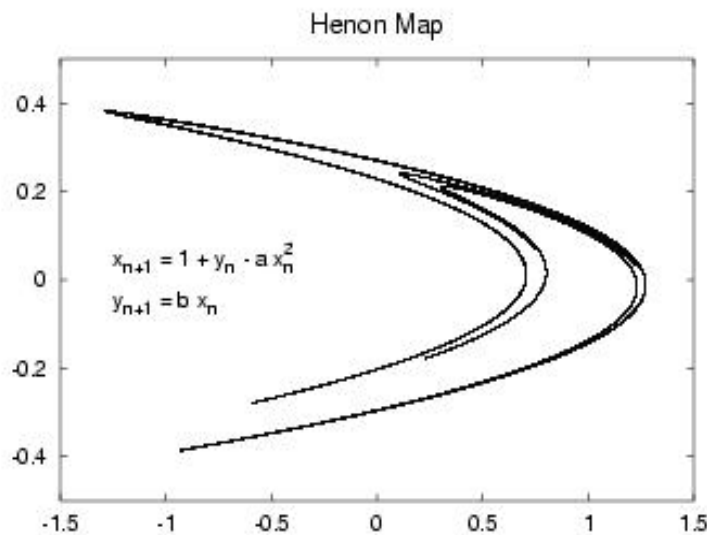
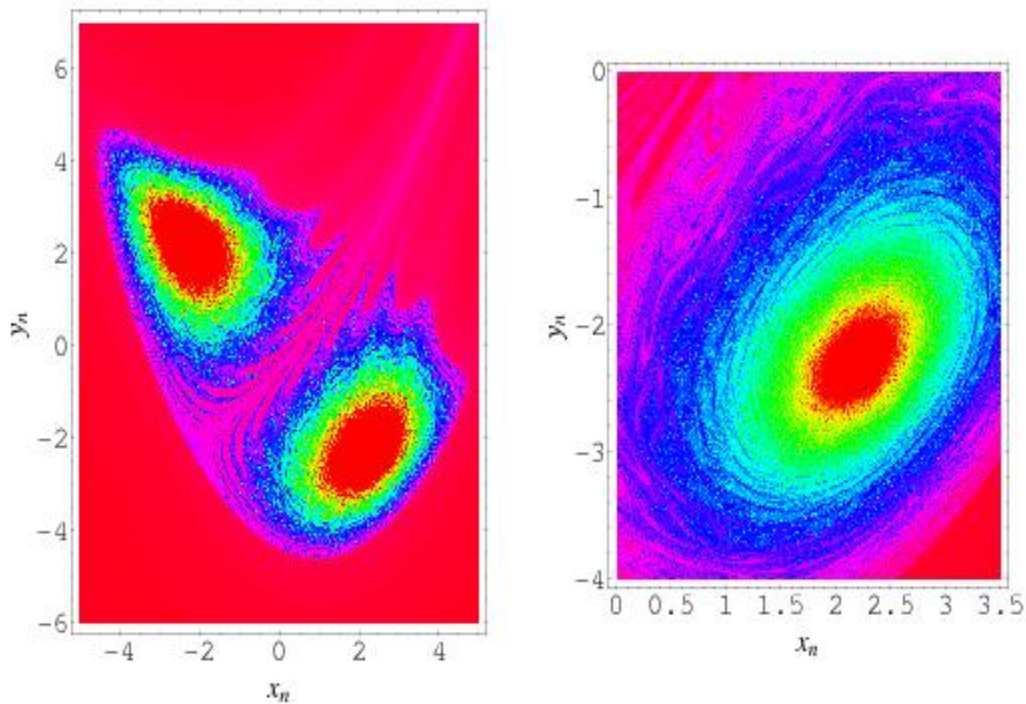


Fig 3.2 2D representation of Henon map

For  $a=1.4$  and  $b=0.5$ , the henon map in Fig 3.2 shows the chaotic behaviour and for the other values of  $a$  and  $b$  it may be choatic,intermittent or obtain to a periodic orbit.The initial parameter  $X_1$  and  $Y_1$  of henon map serves as a symmetric key for chaotic cryptographic system.

The deterministic nature of Henon map, allows the reconstruction of original image at receiver's end from the cipher image during decryption process for the same initial parameters  $X_1$  and  $Y_1$ .Thus, the key sensitivity and encryption algorithm contributes together to avoid all kind of cryptanalysis attacks.[31]



**Fig 3.3 The strange attractor obtained for  $a=1.4$  and  $b=0.3$**

### 3.6 ATTRACTOR

Attractor is a set of variables which evolves in discrete dynamical system. Henon attractor evolves from two parameter mapping given by eq 13:

$$\mathbf{H}_{ab}(\mathbf{x}, \mathbf{y}) = (1 - a\mathbf{x}^2 + \mathbf{y}, b\mathbf{x}) \quad \text{eq(13)}$$

The Hénon attractor is denoted by  $H_A$  and is defined as the set of all points for which the system iterates of every point in a certain quadrilateral  $Q$  surrounding  $H_A$  approach a point in the set. For  $a=1.4$  and  $b=0.3$  Henon map shows the chaotic behaviour, by iterating the eq 14:

$$\mathbf{X}_{n+1} = 1 + \mathbf{Y}_n - a\mathbf{X}_n^2$$

$$\mathbf{Y}_{n+1} = b\mathbf{X}_n \quad \text{eq(14)}$$



**Fig 3.4 Attractor for dynamic system**

This chaotic behaviour is known as Henon attractor is the orbit of the iteration shown in Fig 3.4. The set of variables of attractor evolving in a discrete dynamical system moves dynamically with time and are closely related to each other. They are represented through vector dimension. Thus, attractor is a region in  $n$  dimensional space evolving according to the variable dimensional parameter set. It can be visualized geometrically shown in fig. An attractor can be a curve point or a complicated fractal structure known as strange attractor. [33]

A strange attractor is a dynamic kind of equilibrium as shown in Fig 3.3. The dissimilarity between attractor and a strange attractor is, an attractor graphically represent a state on which the system finally settles down.[34] Whereas strange attractor represents a kind of flight which changes from situation to situations and never gets settles down.

### **3.7 CONFUSION AND DIFFUSION**

In cryptography, for the proper functioning of the secure cipher, Claude Shannon discovered two main properties known as confusion and diffusion in his paper Communication Theory of Secrecy Systems [35] which proves to be very essential for designing a secured encrypted message. Shannon Theory helps in deduction of possible ciphertext attack based on statistical analysis of plaintext. Generally, an intruder uses the prior knowledge of plaintext statistical analysis as a base for ciphertext attack.

According to Shannon's Theory, confusion is defined as building the complex and involved relationship between the ciphertext and the symmetric key and diffusion is defined as dissipating the statistical structure of plaintext over bulk of ciphertext. For implementing this complexity a series of substitutions and permutations technique is practised. [36]

# PROPOSED METHODOLOGY

---

The proposed encryption and authentication method has been divided into three phases:

Phase 1: KEY GENERATION

Phase2: IMAGE ENCRYPTION USING HENON CHAOTIC MAP

Phase3: AUNTHENTICATION OF USER

In the proposed scheme, we first start with obtaining seed value which is obtained from the entropy of the given system and applying DNA cryptography and Sequence Alignment technique. The obtained seed value is shared between two parties and is considered as shared secret key. This seed value serves as an initialization parameter for proposed encryption algorithm. In the proposed encryption method, the images encryption is done by Henon Chaotic System and decryption is done using the same seed value as a decryption key. The user uploads its encrypted data over the cloud, if any other user wants to access this encrypted data. It needs to be get authorized by the authentication server of the cloud. For authenticating, an authentication approach is proposed in which user authentication is carried out through secure key exchange for validating user legal identities and once the authentication is done successfully, the actual symmetric key is given to user through secured channel, through which it can access the encrypted data over the cloud.



## 4.1 KEY GENERATION METHOD

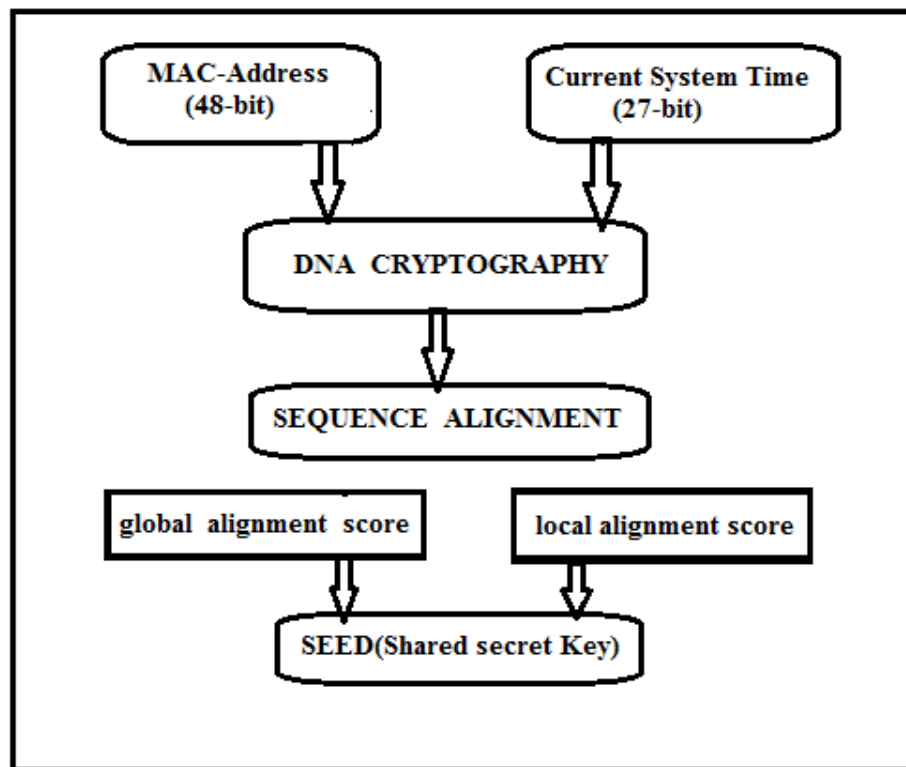


Fig 4.1 Flowchart of Key Generation Method

In this proposed method,

- Step 1: Obtain the MAC address of the current user system who wants to upload the encrypted data over the cloud. The MAC address obtained is in hexadecimal notation which is converted into binary notation to get 48-bit address.
- Step 2: Obtain the current time of the system which consists of three parameters hours, minutes and seconds which is converted into binary notation to get 24-bit message.
- Step 3: Obtained current system entropy i.e the 72-bit message (having MAC address and current system time) is converted into DNA sequence strands on the basis of following rule given by table

S.NO	DNA BASE	CODE	S.NO	DNA BASE	CODE
1	AA	0000	9	GA	1000
2	AC	0001	10	GC	1001
3	AG	0010	11	GG	1010
4	AT	0011	12	GT	1011
5	CA	0100	13	TA	1100
6	CC	0101	14	TC	1101
7	CG	0110	15	TG	1110
8	CT	0100	16	TT	1111

**Table 2 DNA Base Coding**

This method of hiding data in a random DNA sequence is known as DNA Cryptography. Thus for given input 72-bit message we get a 36-bit DNA Sequences which is the compressed form of the original message.

Example of DNA Cryptography:

Suppose, we have a message  $M='10101001'$ , and we need to hide this message in DNA sequence. On the basis of mapping rule specified above we get the corresponding DNA sequence for this given message which is 'GGGC'. Thus, we get a compressed DNA sequence having length 4 just half of the original message.

Step 4: Now, we obtained a random DNA sequence from the large database available having the same length of above obtained DNA sequence and we carried out Sequence Alignment of the known DNA sequence with this random DNA sequence and measure the degree of similarity through obtained local alignment score and global alignment score.

Step 5: We calculate sequence alignment of the known DNA sequence with the random DNA sequence and measure the degree of similarity by calculating the local alignment score and global alignment score.

Here, the global alignment score is obtained by applying Needleman- Wunsch global alignment algorithm and local alignment score is obtained by applying Smith-Waterman local alignment algorithm.

Once, this score matrix is obtained it is considered as seed value which serves as share secret key and used as a initialization parameter in for Henon chaotic system in order to generate pseudo random sequences for the given image.

## 4.2 IMAGE ENCRYPTION USING HENON CHAOTIC MAP

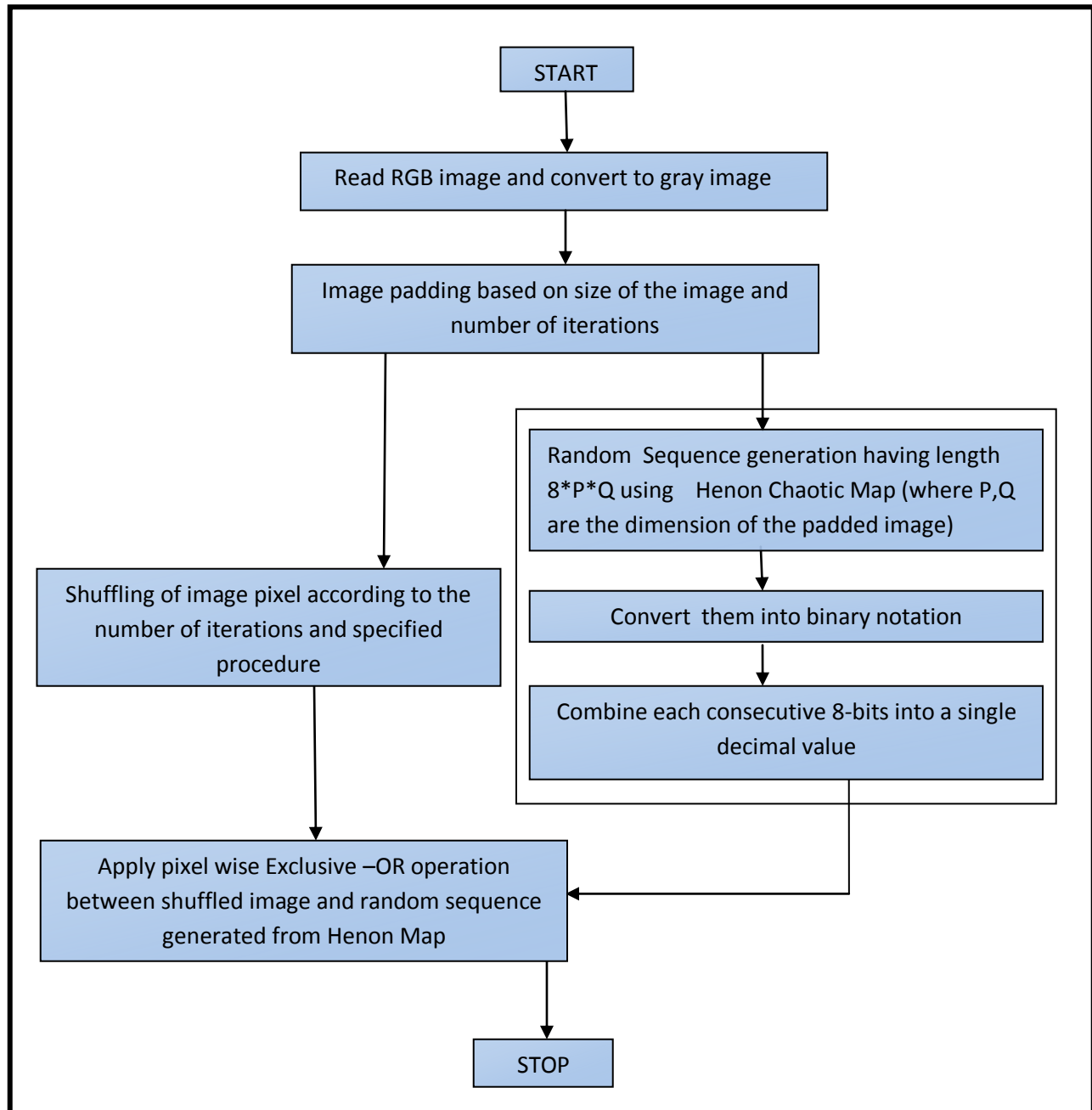


Fig 4.2 Flowchart of Image Encryption Algorithm

### INPUT PARAMETERS:

**I:** Image to be encrypted, having dimension (P\*Q) where P and Q corresponds to number of rows and column respectively.

**X (1):** Initial X-axis parameter for the Henon map

**Y (1):** Initial Y-axis parameter for the Henon map

The X (1) and Y (1) are the shared secret key obtained from the key generation method mentioned above.

### PROPOSED ALGORHITM:

#### Step 1: RGB TO GRAY CONVERSION

The input image is a RGB image which is converted into gray scale image. Where,  $I(x, y)$  represent each pixel intensity at position x and y when the image is converted from RGB to gray image.

#### Step 2: IMAGE PADDING

The obtained gray image is then padded with the zero matrix according to the number of iterations required based upon image size. Pad size(S) is a vector which specifies the dimension and the amount of padding to be done. According to the padding algorithm,

- If the number of rows (P) is not divisible by  $2t$ , where t is number of iterations of image shuffling then we add a zero matrix having pad size  $2t-1$  having dimension equal to  $[1, \text{number of columns}(Q)]$ . Thus, new image size is calculated using eq 15

**New image size = [(number of initial rows (P) +Pad size(S)), (number of initial columns (Q))] eq(15)**

- If the number of columns(Q) is not divisible by  $2^t$ , where  $t$  is number of iterations of image shuffling then we add a zero matrix having pad size  $2^t-1$  having dimension equal to  $[\text{number of rows}(Q), 1]$ . Thus, new image size is calculated using eq 16

**New image size =  $[(\text{number of initial rows (P)}, (\text{number of initial columns (Q)} + \text{Pad size(S)}))]$ eq(16)**

The Padding algorithm is given below,

```

For i = 1 to t
    if  $P/2^t == 0$ 
        No padding required in image
    Else
         $P = P + 2^{t-1}$ ;
         $Q = Q$ ;
        Image new dimension =  $I + [2^t - 1]$  number of padding having size  $[1, Q]$ ;
    if  $Q/2^t == 0$ 
        No padding required in image
    Else
         $P = P$ ;
         $Q = Q + 2^{t-1}$ ;
        Image new dimension =  $I + [2^{t-1}]$  number of padding having size  $[1, P]$ ;
end

```

### Step3: SHUFFLING OF PIXEL BASED UPON NUMBER OF SUCCESSIVE ITERATIONS

Shuffling helps in changing the correlation among the adjacent pixels. With this transformation the image is being apparently randomized and the original image is regained after number of back tracking steps. The shuffling of image depends upon the size of the image. It does the shuffling for the image having even number of rows and columns and if the dimension not matched the extra padding with zero matrix is done. Shuffling of pixel is done in two steps:

For every iteration, a quadrant is sub divided into equal sub-quadrants.

For each  $t^{\text{th}}$  iteration,

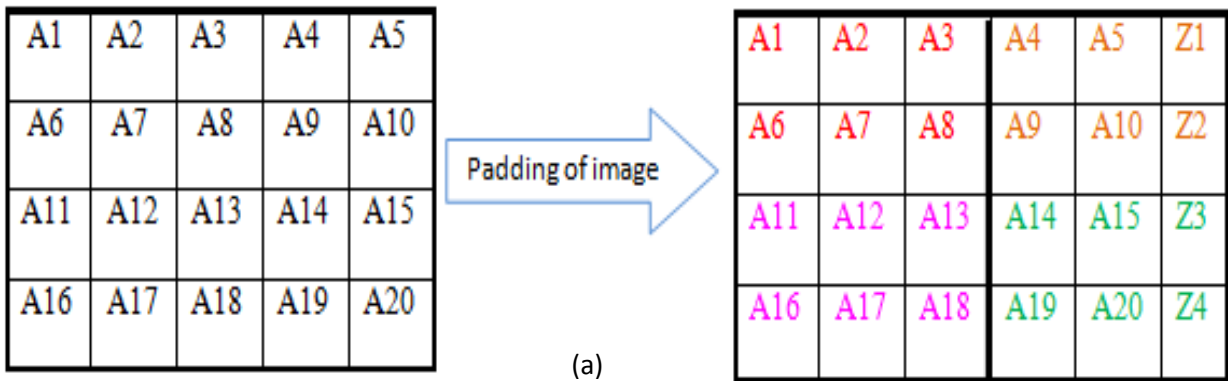
if 't' is odd( $2n-1$ ) then the rotation of quadrant is done in anti-clockwise direction.

If 't' is even( $2n$ ) then the rotation of quadrant is done in clockwise direction.

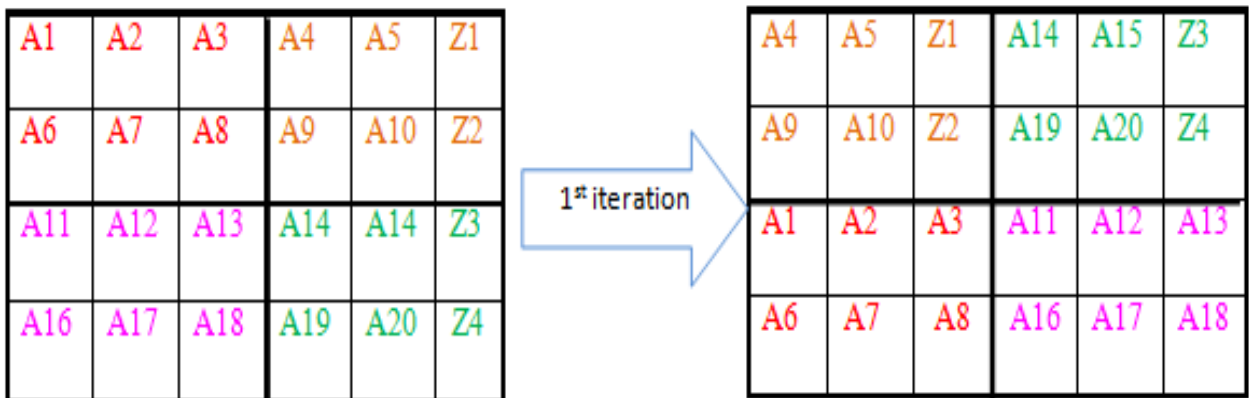
#### **Example:**

To padding and shuffling of pixel is illustrated with an example given below.

Consider an image, having size  $P*Q$ , where  $P=4$  and  $Q=5$ , so total number of pixel in an image is 20. Therefore, it undergoes iterations 3 times, moreover we also see that number of rows  $P=4$  which is even but number of columns  $Q=5$  is not even and thus padding of zero matrix is done having dimension  $[P, 1]$  to the number of columns. the shuffling and padding of image is shown by the Fig 4.3

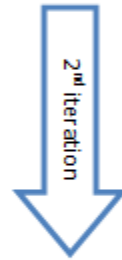


(a)



(b)

A4	A5	Z1	A14	A15	Z3	X1	X2
A9	A10	Z2	A19	A20	Z4	X3	X4
A1	A2	A3	A11	A12	A13	X5	X6
A6	A7	A8	A16	A17	A18	X7	X8



A1	A2	A3	A11	A4	A5	Z1	A14
A6	A7	A8	A16	A9	A10	Z2	A19
A12	A13	X5	X6	A15	Z3	X1	X2
A17	A18	X7	X8	A20	Z4	X3	X4

(c)

**Fig 4.3 (a) Position of pixel in padded image (b) Position of pixel in shuffled image after 1<sup>st</sup> iteration (c) Position of pixel in shuffled image after 2<sup>nd</sup> iteration**

**Step 4: ENCRYPTION OF THE SHUFFLED IMAGE USING HENON CHAOTIC SYSTEM**

The obtained shuffled image is then encrypted using pseudo random binary sequence generated by taking initial values (shared secret key) for Henon map. Henon map is a discrete time dynamical system that exhibits chaotic behaviour. It takes as input initial points  $[X_t, Y_t]$  and maps it into new point through eq 17:

$$X_{t+1} = 1 + Y_t - 1.4X_t^2$$

$$Y_{t+1} = 0.3X_t$$

eq(17)



In this method,

- We input the initial values  $X(1), Y(1)$  for Henon map. These values are obtained in key generation phase which works as a secret symmetric key for Henon map.
- The Henon map works as a key stream generator and it generate a large random key sequence on the basis of input initial value and the length of the sequence depends upon the image size. For the image having input size  $P*Q$  then the length of the Henon sequence will be  $8*P*Q$  obtained by the equation given above.
- According to experimental results, the cut-off point, 0.3992 has been determined in order to balance the sequence. Decimal notation is converted into binary notation as per the specified threshold rule given in eq18

$$C_k = 0; \text{ if } Z_k \leq 0.3992$$

$$1; \text{ if } Z_k > 0.3992 \quad \text{eq (18)}$$

- Reduction of obtained Henon sequence is done by combining each consecutive 8-bits and converting them into decimal notation.

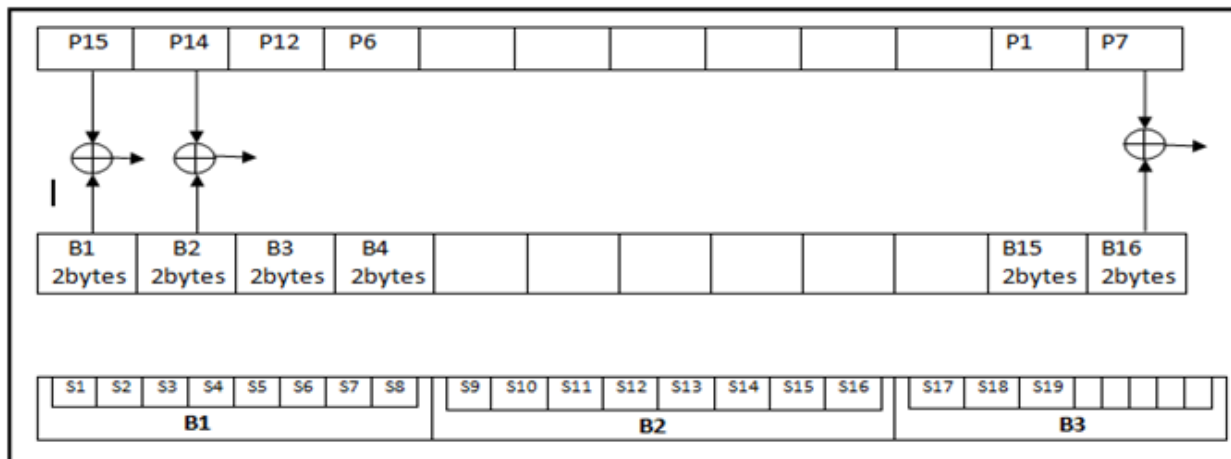
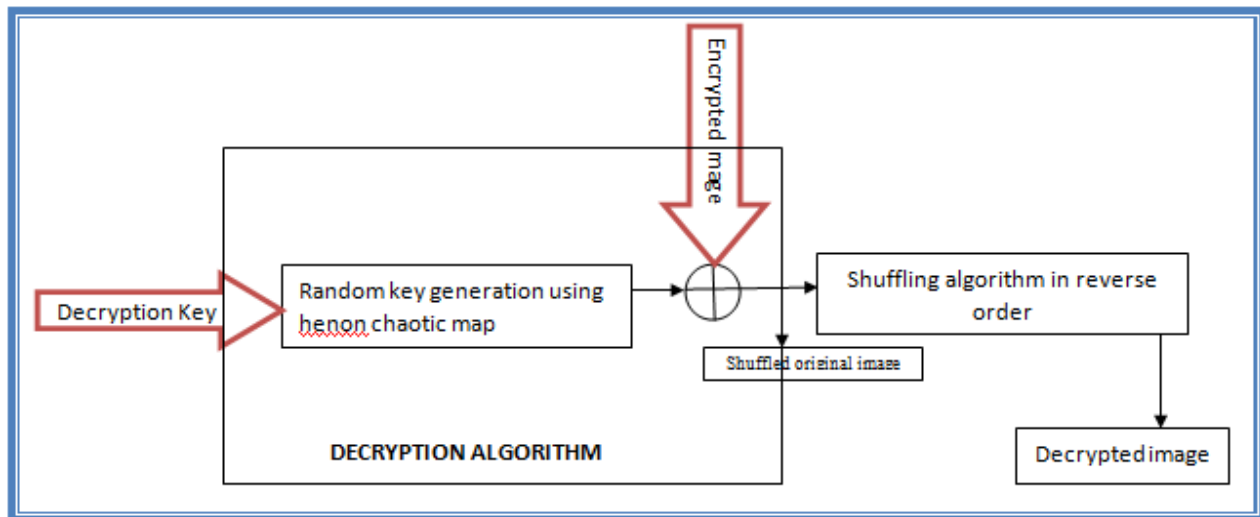


Fig 4.4 Encryption with Byte Sequence

- This is the final step in which bitwise Exclusive-OR operation is done between the pixel of shuffled image and the sequence generated from the Henon map shown in Fig 4.4. Here the confusion is done by the shuffling of pixel on different permutation and diffusion is done by the encryption technique.

#### Step 5: DECRYPTION OF ENCRYPTED IMAGE

The encrypted image is uploaded on the cloud, along with the authentication details i.e. the MAC address and current time of the system. If the other user wants to decrypt it first needs to get the key i.e. the initialization parameter of the Henon Map. Once, the initialization parameter is obtained, the decryption is done to get the shuffled image. This shuffled image is rearranged in the same order by back tracking the encryption algorithm in order to obtain the original image. The flowchart for the decryption process is shown in Fig 4.5



**Fig 4.5 Decryption Process**

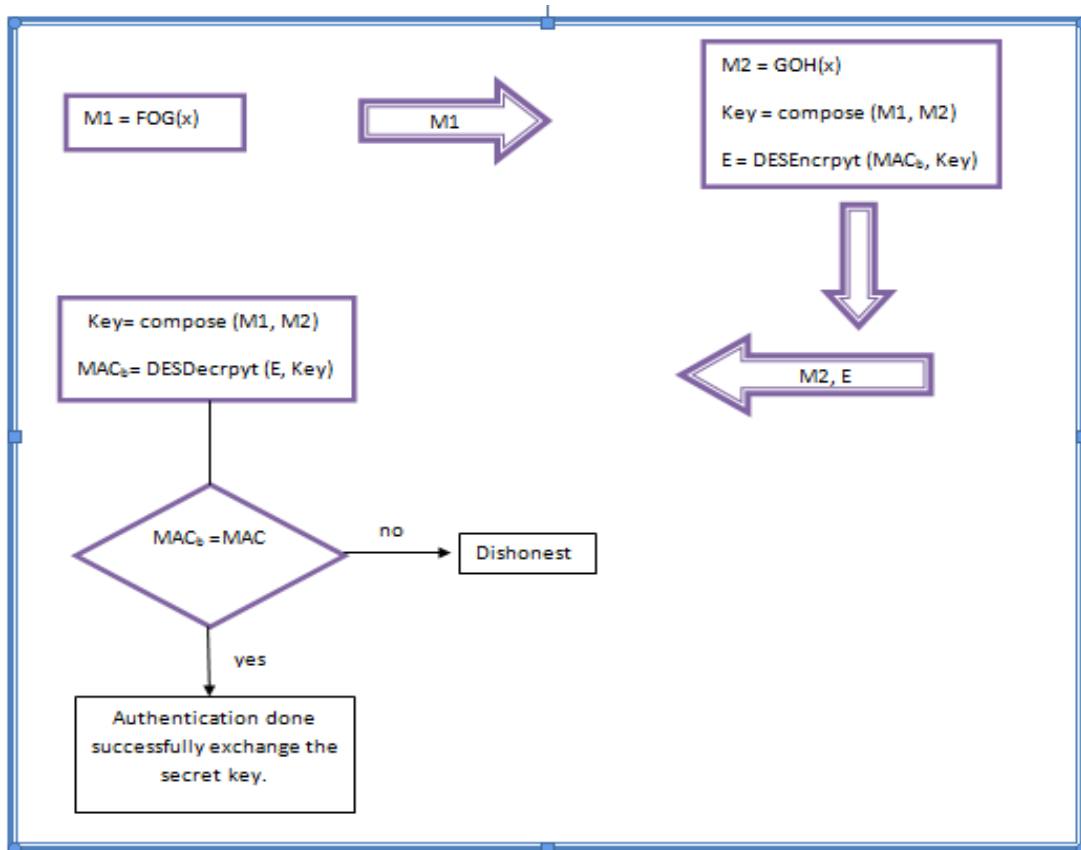
### 4.3 AUTHENTICATION OF USER

In this phase, we proposed a novel user authentication approach and secure key exchange in order to validate the legitimate identity of user. We proposed to use a modified version of the two-server model in which there are two servers one is authentication server and another is database server. The authentication server is responsible for validating the user request and once the user get authorized it get access to encrypted data stored on database server. The database server stored the encrypted data and list of legitimate user who can access this data. The list consists of MAC address of those users to whom the access is granted. The user authentication and encryption steps are explained below:

Suppose, there are two clients USER A and USER B, authentication server (AS) and database server(DS).

USER A wants to upload a file (image file .jpeg) on cloud server; it first generates the key using key generation algorithm mentioned above and encrypts the file with this key using Henon Chaotic System encryption algorithm described above. It also provides the list of legitimate user to authentication server with whom USER A wants to share this encrypted file.

Now, if USER B wants to read this file it first needs to be get authorized. The authorization is done on the basis of MAC address. If USER B MAC address matches with the specified MAC address in the legitimate user list stored over the authentication server (AS), then USER B is successfully validated and can get access to the encrypted file stored on database server (DS).



**Fig 4.6 Flowchart of user authentication and key exchange algorithm**

The User Authentication and Key Exchange Algorithm is follows:

Step 1: The authentication server (AS) chooses a secret random DNA sequence from the available database and converts it into a polynomial function  $F(x)$ .

Step 2: USER B also chooses a secret random DNA sequence and converts it into a polynomial function  $H(x)$ .

Step 3: Both the parties, agreed upon common random DNA Sequence and converts it into a polynomial function  $G(x)$ , which is public and known to both the parties.

Step 4: The AS calculates composition of  $F(x)$  and  $G(x)$  and send it to USER B Message  $(M1) = \{FOG(x), G(x)\}$

$$\mathbf{M1=FOG(x) = compose [F(x), G(x)]} \quad \text{eq (19)}$$

Step 5: USER B also calculates the composition of  $G(x)$  and  $H(x)$

$$\mathbf{M2=GOH(x) = compose [G(x), H(x)]} \quad \text{eq (20)}$$

Step 6: Now, the USER B calculates the symmetric key with the help of  $M1$  and  $M2$

$$\mathbf{FOGOH(x) = compose [FOG(x), GOH(x)]} \quad \text{eq (21)}$$

$$\mathbf{K_{ss} = FOGO H (x')}; \text{ where } x' \text{ is the sum of all the coefficient of } x \text{ in the polynomial } FOGO H(x)$$

and with the help of this symmetric key it encrypts its MAC-address(48-bit) using DES symmetric key encryption algorithm.

$$\mathbf{H'=DES-Encrypt [MAC-address, K_{ss}]} \quad \text{eq (22)}$$

Thus, USER B sends to authentication server (AS),  $H'$  and  $M2$ .

Step 7: The Authentication server receives  $M2$  and calculates the symmetric key  $K_{ss}$  with the help of  $M1$  and  $M2$  and with this key it decrypts the received encrypted message  $H'$  using DES decryption algorithm.

$$\mathbf{MAC'=DES-decrypt [H', K_{ss}]} \quad \text{eq(23)}$$

If the obtained MAC-address ( $MAC'$ ) matches with the MAC address entered in the legitimate user list stored over the authentication server. Then the USER B is successfully validated and gets authorized by the AS.

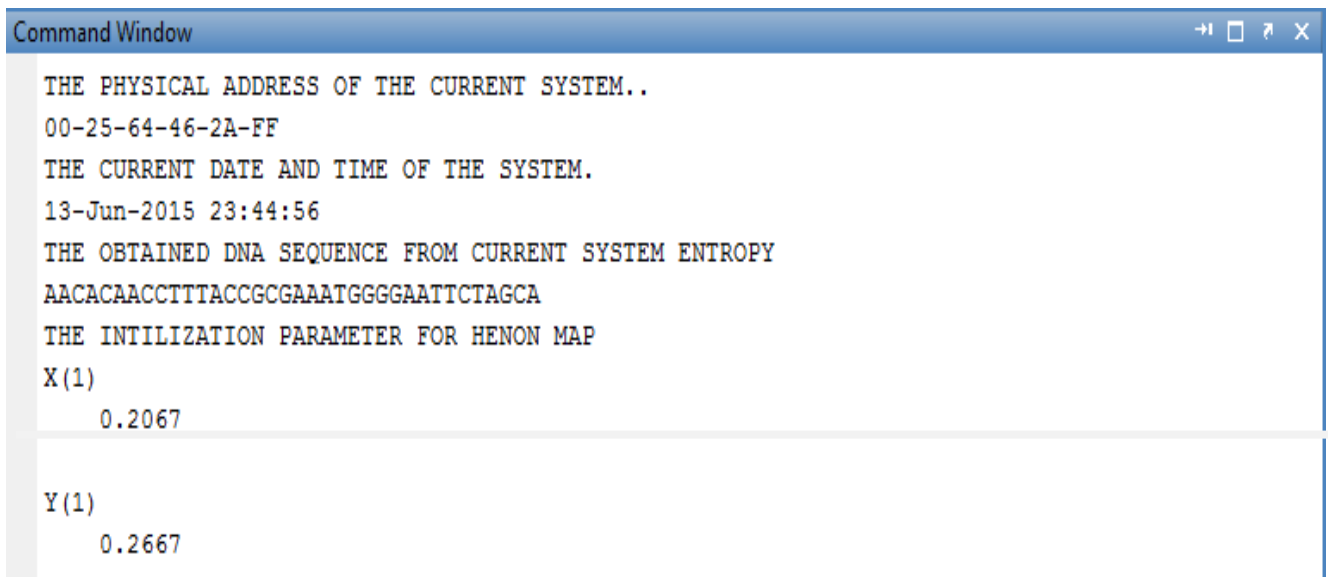
Step 8: Once the USER B gets authorized successfully, it request AS for the key and AS in turn request DS the obtained key is securely given to the USER B through secured channel. Through this key, it can decrypt the encrypted image uploaded over the cloud.

# EXPERIMENTAL RESULTS & SECURITY ANALYSIS

The following system configuration has been used while conducting the experiments:

- Processor: Intel Core i3
- Main Memory: 4 GB
- Hard Disk Capacity: 512 GB
- Software Used: MATLAB R2010a, NIST TEST SUITE

## 5.1 KEY GENERATION METHOD

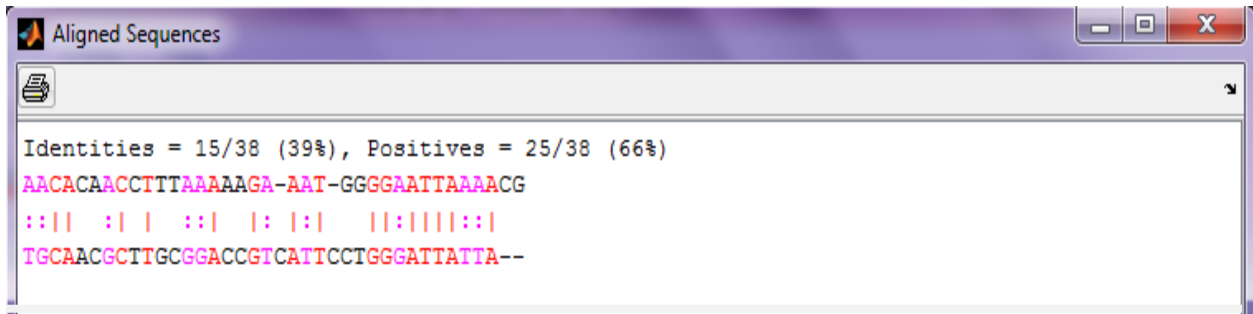


```
Command Window
THE PHYSICAL ADDRESS OF THE CURRENT SYSTEM..
00-25-64-46-2A-FF
THE CURRENT DATE AND TIME OF THE SYSTEM.
13-Jun-2015 23:44:56
THE OBTAINED DNA SEQUENCE FROM CURRENT SYSTEM ENTROPY
AACACAACCTTTACCGCGAAATGGGGAATTCTAGCA
THE INTILIZATION PARAMETER FOR HENON MAP
X(1)
    0.2067
Y(1)
    0.2667
```

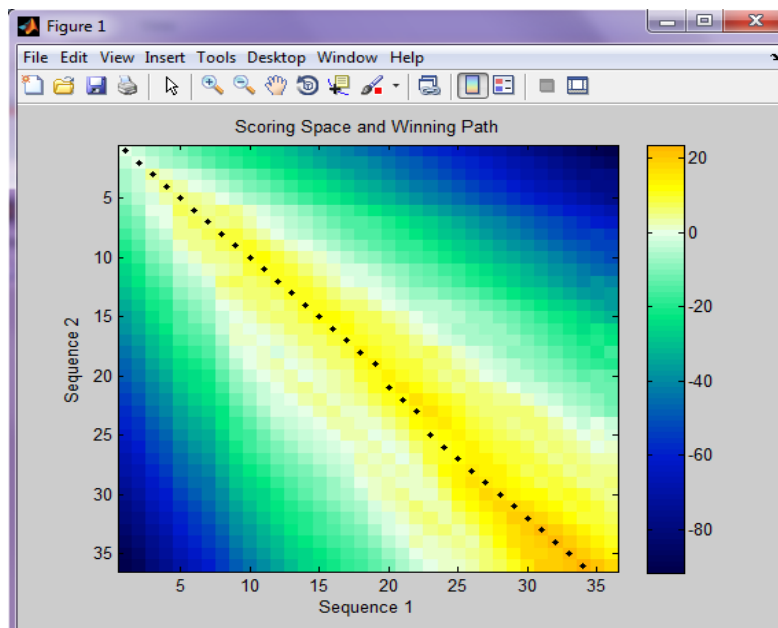
**Fig 5.1** The random seed (secret key) value obtained from DNA sequence and MAC address using sequence alignment.

The sequence alignment of sequences with the random DNA is shown below:

### 5.1.1 GLOBAL ALIGNMENT AND WINNING PATH



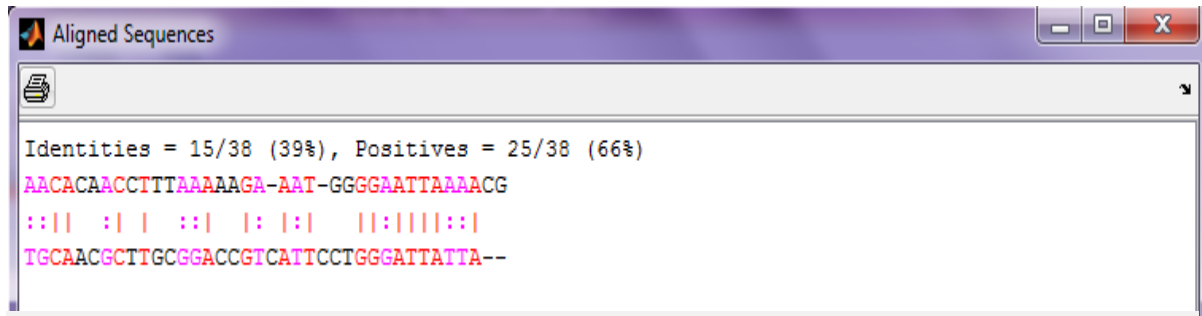
(a)



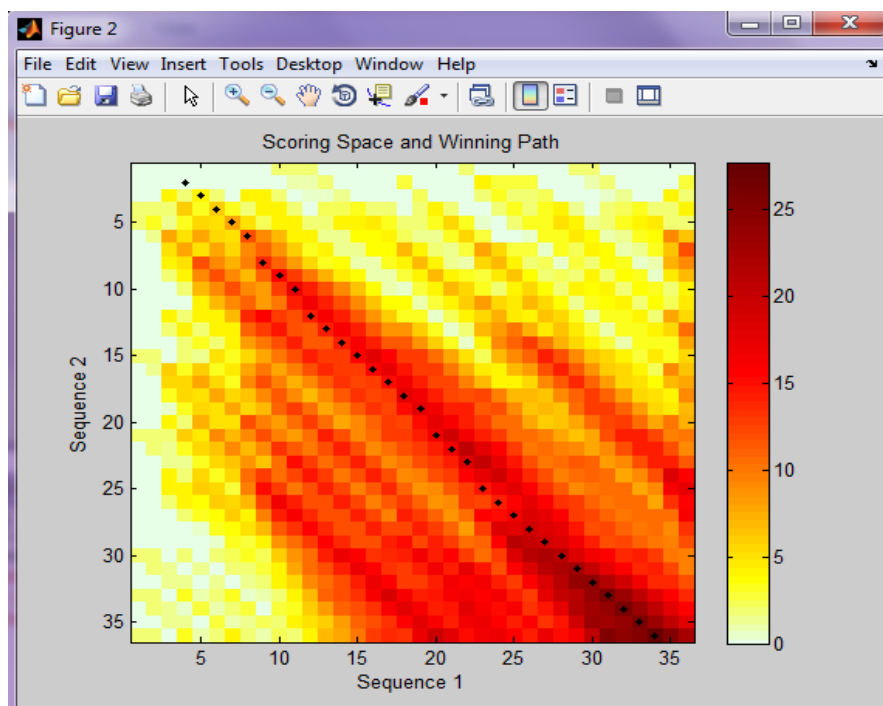
(b)

**Fig 5.2 (a) Global alignment of random DNA sequence and DNA sequence obtained from system entropy (b) Winning Path Matrix**

## 5.1.2 LOCAL ALIGNMENT AND WINNING PATH



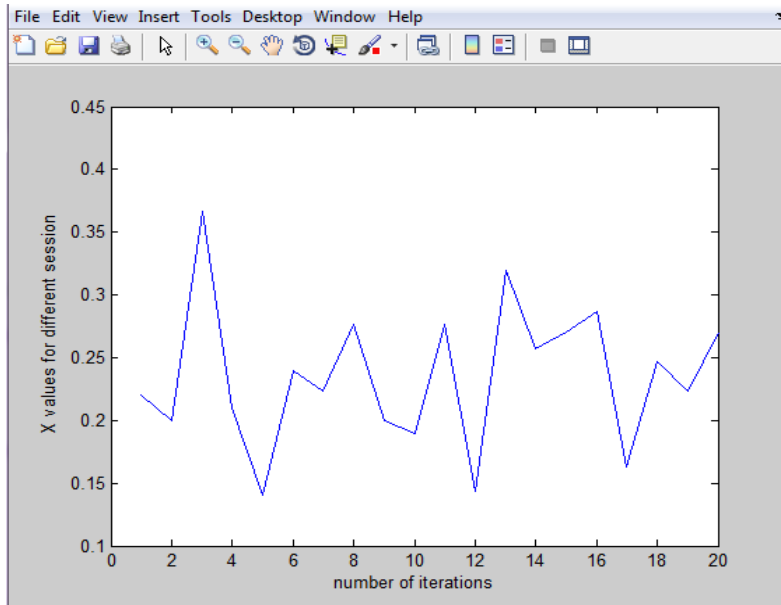
(a)



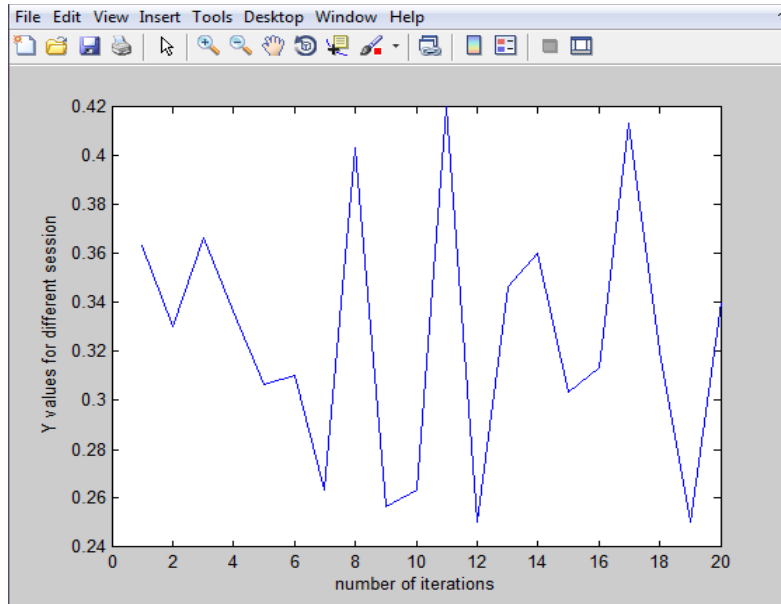
(b)

Fig 5.3 (a) Local alignment of random DNA sequence and DNA sequence obtained from system entropy (b) Winning Path Matrix





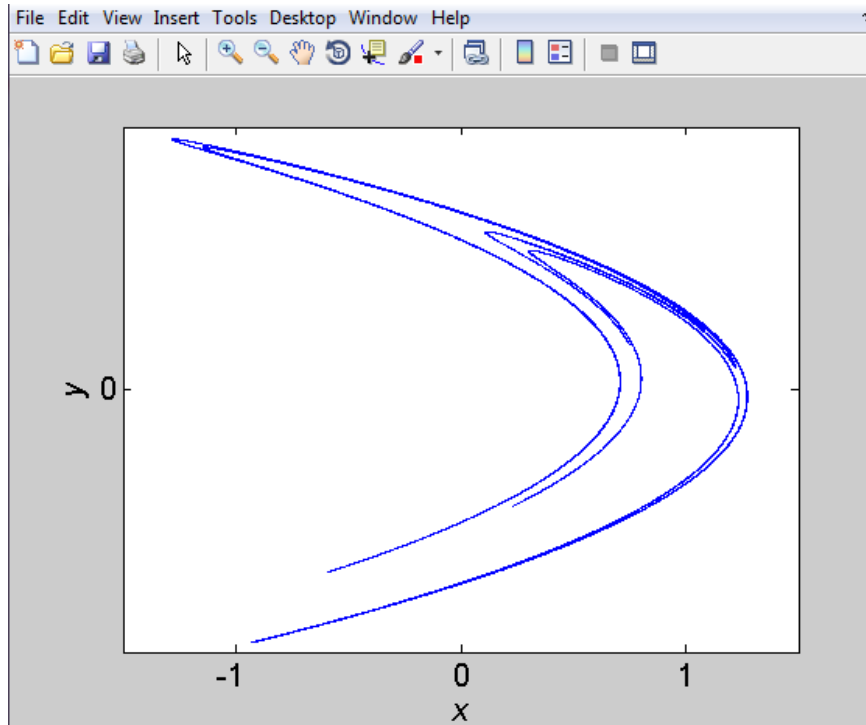
**Fig 5.4 The randomness walk for X value**



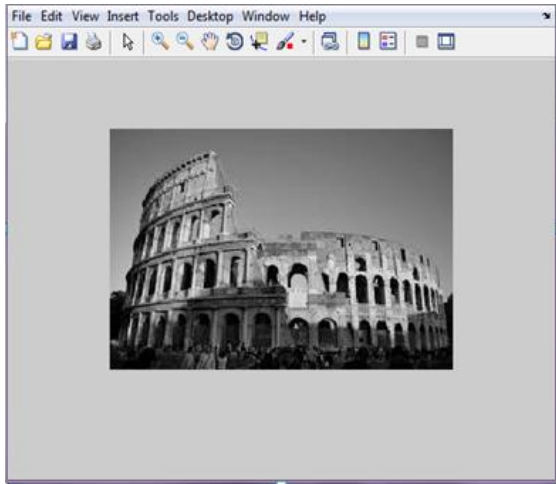
**Fig 5.5 The randomness walk for Y value**

## 5.2 ENCRYPTION AND DECRYPTION OF THE IMAGE

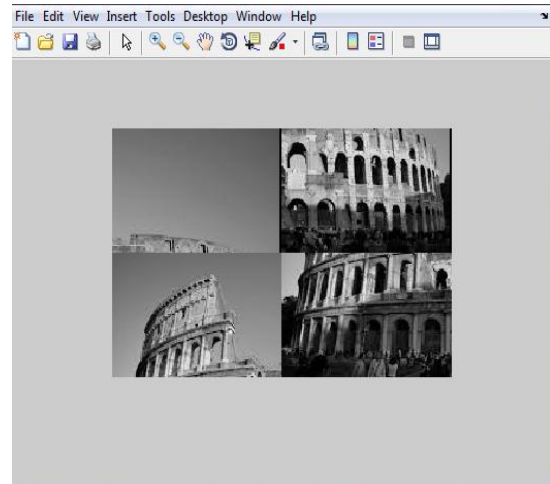
Here, we have taken the test image of size 302\*256 as shown in figure 5.7. The plot of Henon map for the initial parameter  $a=1.4$  and  $b=0.3$  to obtain a chaotic system is shown in figure 5.6. The initialization parameter  $X(1), Y(1)$  for the Henon map are obtained from global alignment score value and local alignment score value.



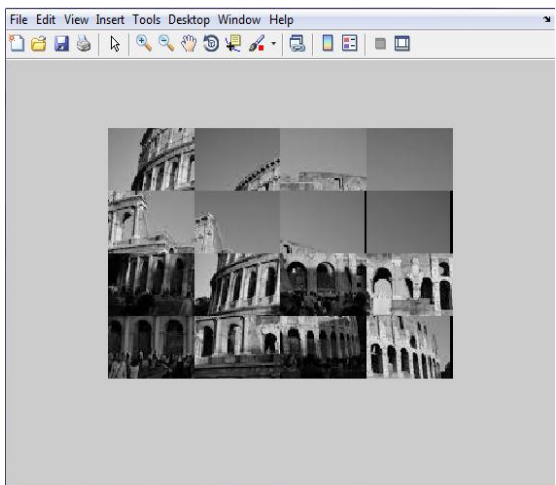
**Fig 5.6 Plot of Henon map**



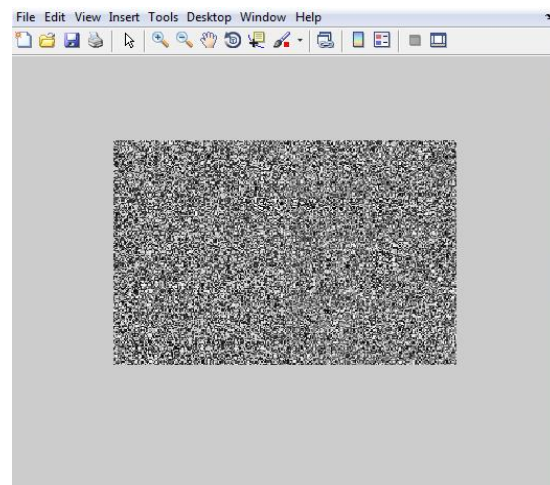
(a)



(b)

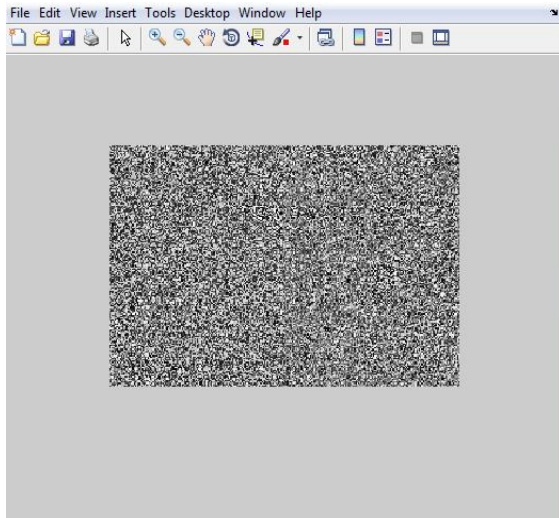


(c)

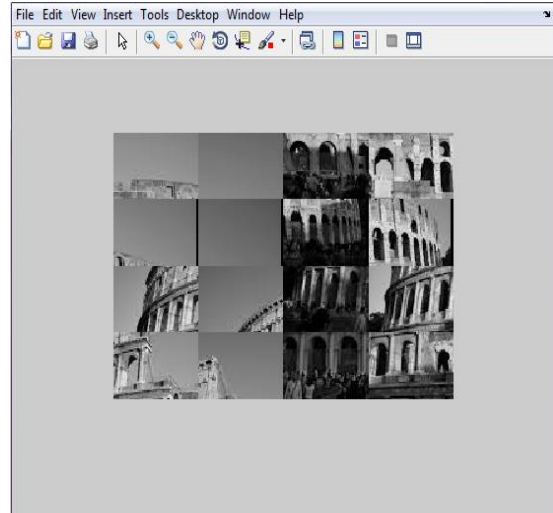


(d)

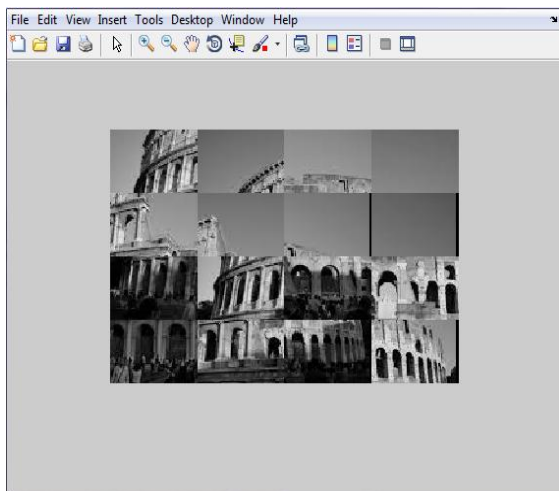
**Fig 5.7 Encryption by Henon chaotic system. (a) Original image (b) Shuffled image after 1<sup>st</sup> iteration [anticlockwise] (c) Shuffled image after second iteration [clockwise] (d) Cipher image**



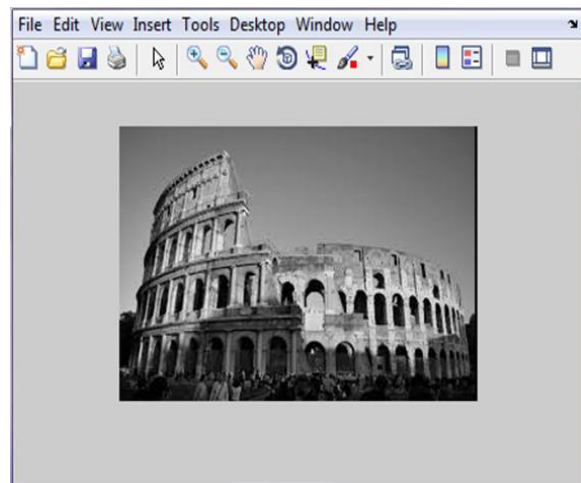
(a)



(b)



(c)



(d)

**Fig 5.8 Decryption by Henon chaotic system. (a) Cipher image (b) Shuffled image after second iteration [anti clockwise] (c) Shuffled image after 2<sup>nd</sup> iteration [clockwise] (d) Original image**

## 5.3 AUTHENTICATION OF USER

```

MATLAB 7.10.0 (R2010a)
File Edit Debug Parallel Desktop Window Help
Current Folder: F:\
Shortcuts How to Add What's New
Command Window
the system mac address
D4-BE-D9-46-B9-53
THE CURRENT DATE AND TIME OF THE SYSTEM.
22-Jun-2015 05:13:00
polynomial function and key value obtained at authentication server side
480*x + 72*(6*x^2 + 20*x + 2)^2 + 144*x^2 + 3*(40*x + 6*(6*x^2 + 20*x + 2)^2 + 12*x^2 + 18)^2 + 227

6.7302e+067

polynomial function obtained at user B side
480*x + 72*(6*x^2 + 20*x + 2)^2 + 144*x^2 + 3*(40*x + 6*(6*x^2 + 20*x + 2)^2 + 12*x^2 + 18)^2 + 227

6.7302e+067

yes
the plaintext i.e mac address to be encrypted is
Columns 1 through 27

1 1 1 0 1 0 0 0 1 0 1 1 0 1 0 0 0 1 0 0 0 1 0 0 1 1 1

Columns 28 through 54

0 1 1 0 0 0 1 1 0 1 0 0 1 1 1 0 1 0 0 1 1 0 1 0 1 0 1

Columns 55 through 64

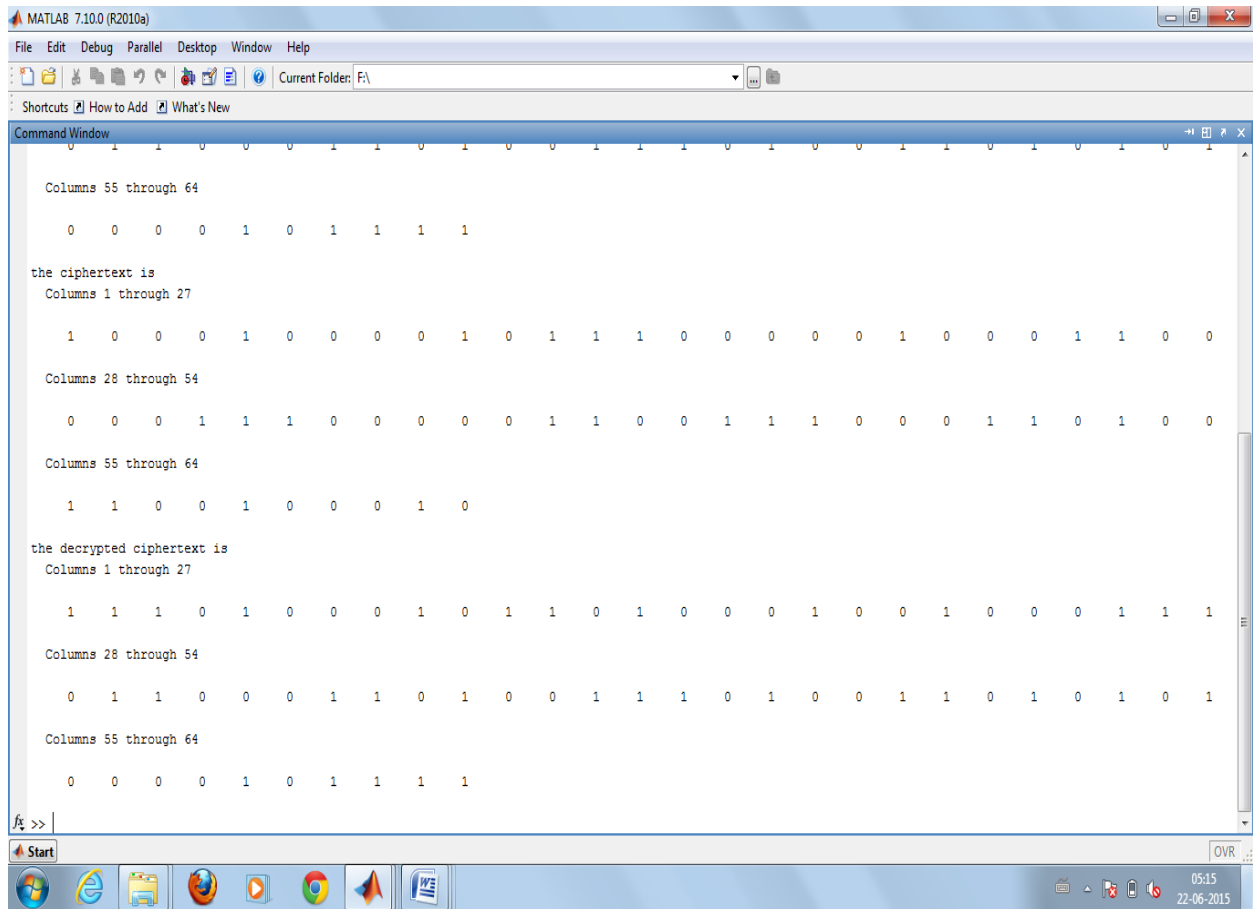
0 0 0 0 1 0 1 1 1 1

the ciphertext is
Columns 1 through 27

1 0 0 0 1 0 0 0 0 1 0 1 1 1 0 0 0 0 1 0 0 0 1 1 0 0

```

(a)



(b)

**Fig 5.9 (a) Showing the key exchange between the user and authentication server (b) encryption and decryption of MAC address and current system using shared symmetric key obtained in key exchange algorithm.**

## 5.4 STATISTICAL ANALYSIS:

**5.4.1 Randomness test:** A random seed is a vector or a number which helps in initialization of random number generator or a pseudo random number generator. The selection of an appropriate random seed is important in the area of computer security. Generally, the generation of random seed is done from the state of the computer system (such as the time). When the same random seed is shared deliberately, it becomes a secret key. Both the random number generated and pseudorandom numbers generated for cryptographic applications should be unpredictable. There are two types of predictability:

**Forward Predictability:** For a PRNG, if the seed value is unknown, and in spite of having the knowledge of previous output number in the sequence, the probability of guessing the next random number is unpredictable. In our proposed method, the forward predictability is ensure as guessing the next pseudo random binary sequence without knowing current the random DNA Sequence(seed value) is tedious task.

**Backward Predictability:** For a PRNG, in spite of having the knowledge of any generated values, the possibility of guessing the seed should be negligible. Every element of the generated sequence should appear to be an independent outcome having probability  $1/2(0 \text{ or } 1)$ .The backward predictability is also ensured in our proposed method. If one knows the pseudo random sequence i.e. the long stream of binary sequence generated, the feasibility of determining the exact seed value i.e. the random DNA sequence is negligible.

For analysing the randomness of the key value we perform certain test defined by NIST Test Suite. We have carried out test using a random DNA sequence of length 10 and the pseudo random binary sequence generated from it is having length 300.( $n=300$ , where  $n$  is length of entire bit string under test) and  $\alpha=0.001$ .

**P-value:** It gives the probability measure that a perfect random number generator would have produced a sequence less random than the sequence that was tested, given the kind of non-randomness assessed by the test. Ideal P-value for randomness is one and for non-randomness is zero.

Generally, Significance level ( $\alpha$ ) = 0.01

Inference from P-value:

Sequence appears to be random ; if P-value  $\geq \alpha$

Sequence appears to be non-random ; if P-value  $< \alpha$

### **Frequency (Monobit) Test**

This test aims to determine that the count of 1 and 0 in a sequence are approximately the same as in a pure random sequence. It measures the nearness of the fraction of 1 to  $\frac{1}{2}$ , i.e., the number of 1 and 0 in a sequence should be approximately equal

### **Frequency Block Test**

This test aims to measure whether the frequency of 1 in a K-bit block is approximately  $K/2$ , as would be expected under a hypothesis of randomness.

### **Runs Test**

This test aims to measure whether the count of ones and zeros of different length is same as assumed for a random sequence. It gives the measure of the oscillation between 1 and 0 is slow or fast.

### **Longest Runs of One-Test**

The test aims to determine whether the count of the longest run of 1 in the query sequence is consistent when compared to the length of the longest run of 1 defined for pure random sequence.



### **Approximate Entropy Test**

The aim of this test is to compare the frequency of overlapping blocks of two consecutive/adjacent lengths ( $m$  and  $m+1$ ) against the expected result for a random sequence.

### **Cumulative Sums (Cusums) Test**

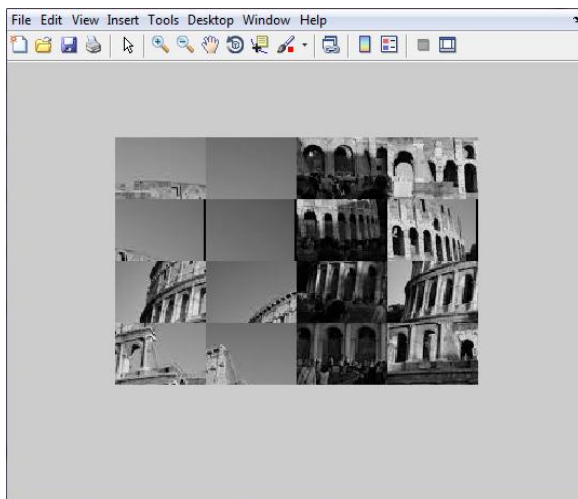
The aim of this test is to calculate the closeness of the cumulative sum of the partial sequences obtained in the query sequence is too large or too small relative to the defined behaviour of that cumulative sum for pure random sequences. The random walk of the sequence is defined for this cumulative sum.

<b>S.NO</b>	<b>NIST TEST SUITE</b>	<b>P-Value (n=300,M=5)</b>
1	Frequency (Monobit) Test	0.151829
2	Frequency Block Test	0.603493
3	Runs Test	0.604353
4	Longest Runs of One-Test	0.120032
5	Approximate Entropy Test	0.365422
6	Cumulative Sums (Cusums) Test	0.231465 (forward)

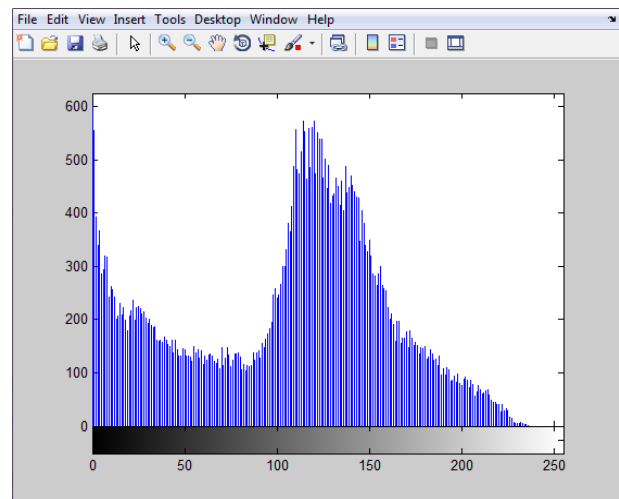
**Table 3 Results of NIST Test Suite**

### 5.4.2 Key Sensitivity Test:

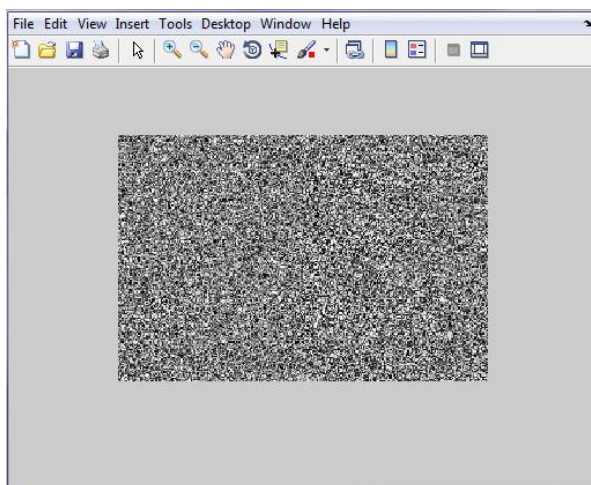
In order to have secure encryption, the key should have large space key size and can resist all kind of Brute force Attack. Key should be sensitive i.e. a minor change in the key value results a major deviation in the decrypted image compared to the original image. For e.g. On changing the initial shared secret key value i.e.  $X(1)=0.034$  to  $X'(1)=0.0333333$  and  $Y(1)=0.076$  to  $Y'(1)=0.0744444$ , it won't be possible to regain the original image with this decrypted key at the receiver's end.



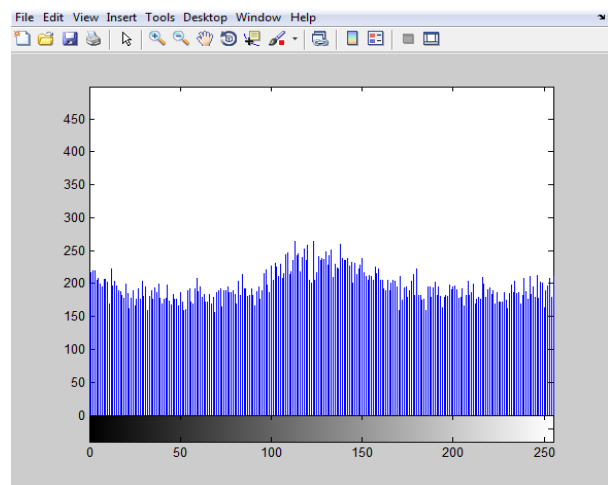
(a)



(b)



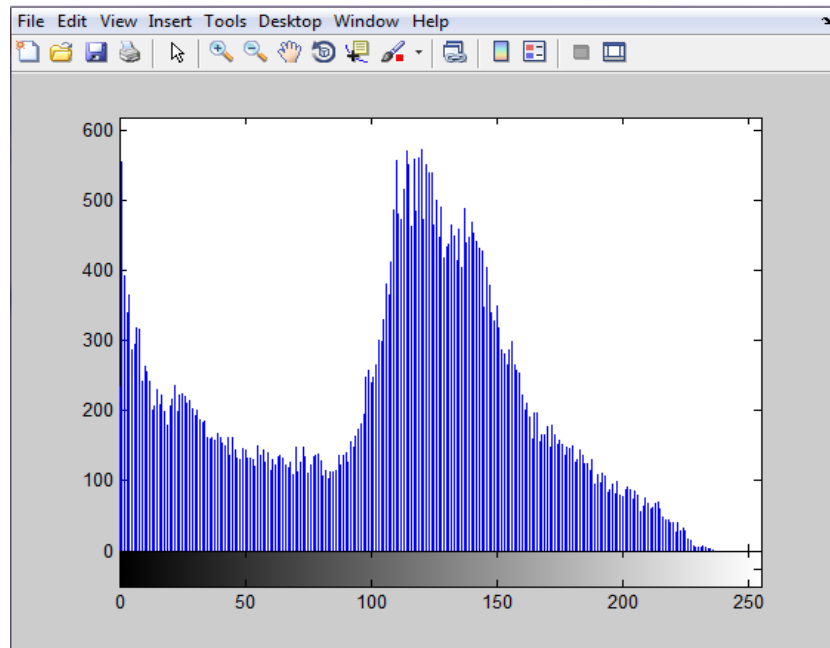
(c)



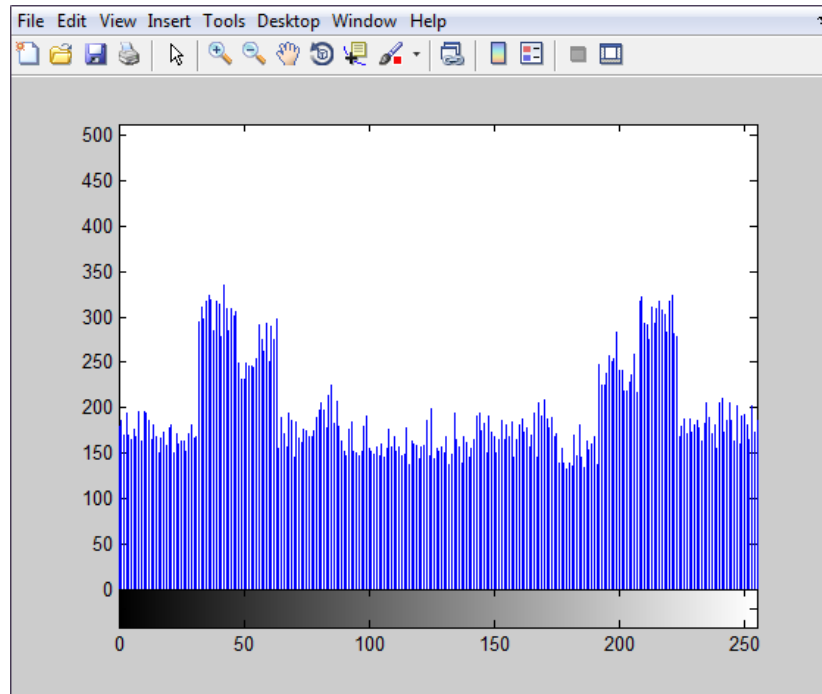
(d)

**Fig 5.10 (a) The decrypted image before shuffling iteration (b) histogram of the decrypted image (c) The decrypted image with slight variation in the key (d) histogram of the decrypted image with slight variation in the key**

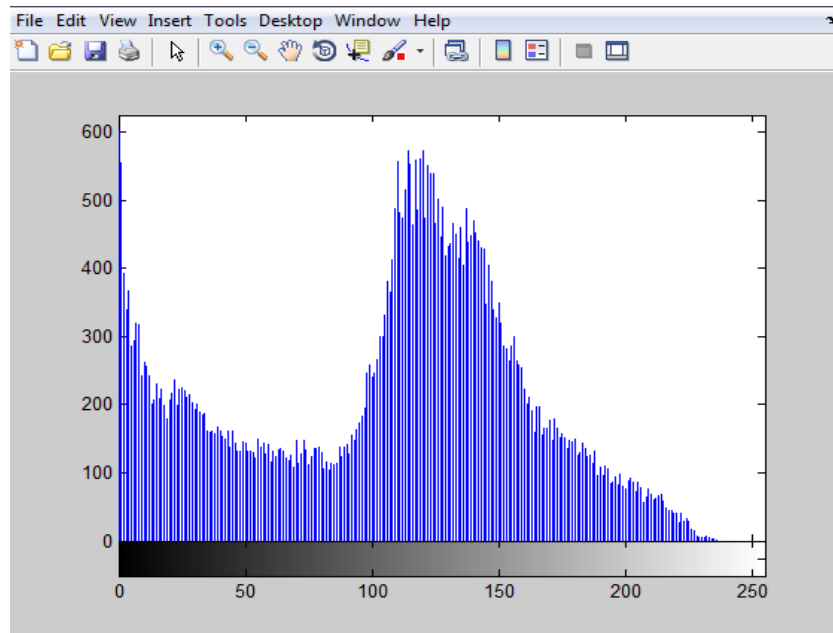
**5.4.3 Histogram Analysis:** Histogram of an image gives the graphical representation of an image in the form of pixel intensity values. For an 8-bit grayscale image, there are total 0-255 different possible intensities values. Histogram counts the number of pixels for each intensity value and plots them.



(a)



(b)



(c)

**Fig 5.11 Histogram analysis: (a) histogram of original image (b) Histogram of encrypted image (c) Histogram of decrypted image**

**5.4.4 Information entropy Analysis:** In the encryption system, the information entropy is defined as the degree of uncertainties in the encryption system. It gives the measure of the Effectiveness of the image encryption algorithm. Entropy is the statistical measure of randomness that can be used to characterize the texture of the input image.

Entropy is defined as:

$$H = -\sum (p_i \cdot \log_2(p_i)) \quad \text{eq (24)}$$

For an encrypted image, the ideal entropy value is 8, which corresponds to a random source. Practically, information entropy is less diverse than the ideal one. The information entropy for the given image is shown in Table

	Original Image	After First Iteration	After Second Iteration	Encrypted Image
ENTROPY	7.5623	7.5579	7.5579	7.9564

**Table 4 Information Entropy Analysis**

Thus, we can see that entropy for encrypted image is 7.9564, which is approximate to the ideal entropy value.

## 5.5 SECURITY ANALYSIS

**5.5.1 Man in the middle attack:** In this attack, the attacker aims at circumventing the process of mutual authentication by forging the identity of each endpoint up to the expected satisfaction level from the legitimate at other end as shown in Fig 5.12.

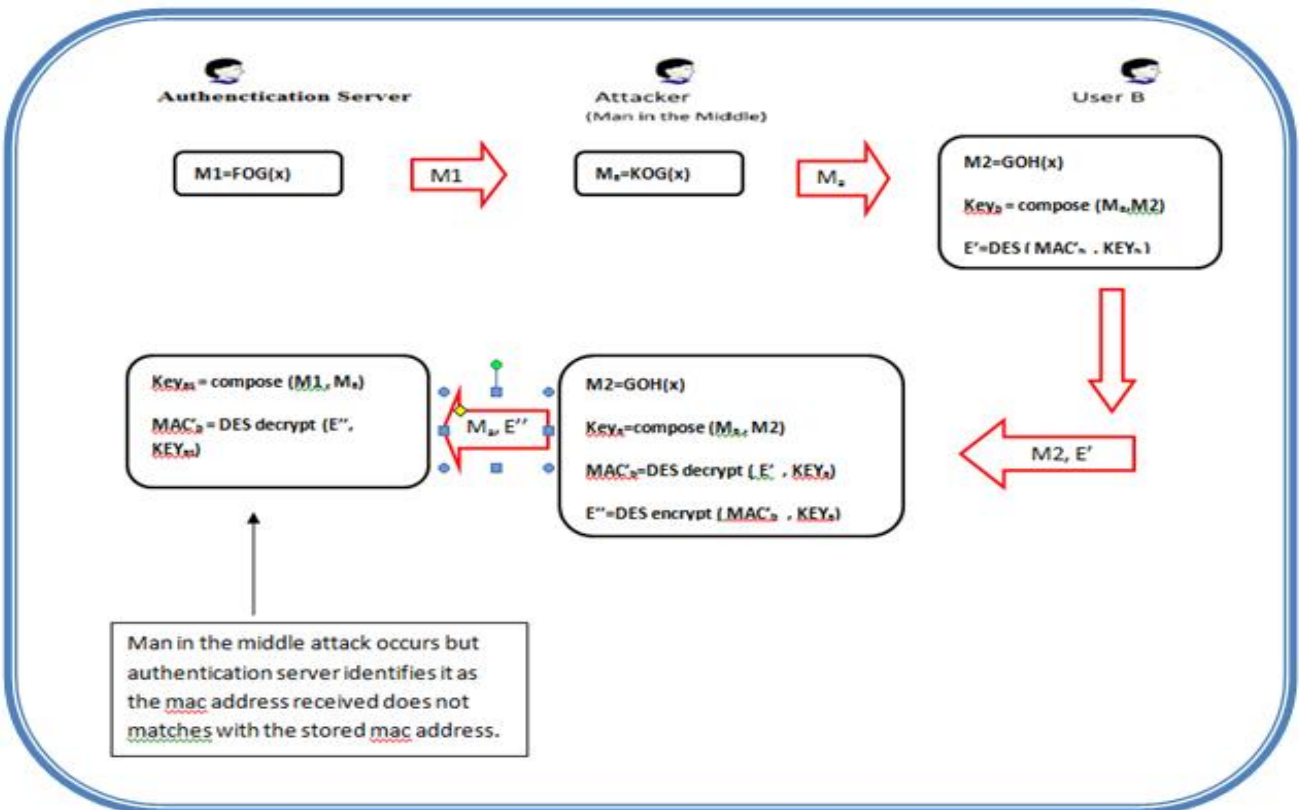


Fig 5.12 Showing Man-in the middle Attack

In this attack, the attacker tries to calculate the symmetric key by taking the partial secret from USER B and from AS, and encrypt the MAC address of USER B with this key, which is detected by authentication server as the decryption with the attacker key gives a different MAC address which does not matches with the existing MAC address in the legitimate user list.

Since, the  $\text{Key}_{(as)} \neq \text{Key}_{(attacker)} = \text{Key}_b$  because of this man in the middle attack was noticed by the authentication server and prevented.

**5.5.2 Brute Force Attack:** For both the encryption algorithm, the probability of guessing all possible combinations of the key is very least and tedious task. The probability of accurately guessing the random DNA sequence from 163 million targets from any sort of DNA database is unachievable. Moreover, guessing the exact random DNA sequence from the polynomial  $G(x)$  and  $FOG(x)$  (both are higher degree of polynomials) is a tedious task. For the authentication algorithm, we have used a key size of 64-bit and DES encryption algorithm is used, we can ensure more security if the key size is increased to 128-bit and AES is used for the encryption of MAC address.

**5.5.3 Replay Attack:** For each authentication session, only one time session key is generated from the random DNA Sequence through which the polynomial function is generated. Both AS and USER B generates the random polynomial function for a single authentication session which gets expired with the session. Thus, the key values changes every time for each authentication session. Hence, our scheme is strong against replay attack.

# CONCLUSION AND FUTURE WORK

---

## 6.1 CONCLUSION

This thesis work proposed an efficient cryptographic and authentication framework for cloud computing with many security features such as mutual authentication, secure key exchange, data isolation and data integrity. In this work, three approaches were proposed they are key generation, image encryption using Henon Chaotic Systems, user authentication. The encryption algorithm is based on chaotic systems which are known for randomness and unpredictable behaviour, so it is highly secured. The confusion is enhanced by shuffling of pixel and diffusion is enhanced through byte sequence generated using Henon chaotic system. To illustrate the efficiency of the cryptographic method statistical analysis like key sensitivity test, histogram analysis, information entropy analysis is done which gives a judgemental result of encryption cryptosystem. Moreover, we perform the NIST test suite like frequency monobit test, frequency test within a block, runs test, longest runs of one's test these test helps in determining the different types of randomness exists in a sequence. The proposed user authentication framework can resist many security attack like replay attack, man-in the middle attack and brute force attack. Thus, the evaluation of this proposed methodology shows that proposed model is able to fulfil the research objectives and helps in reducing the cloud computing security concerns such as authentication and data integrity.

## 6.2 FUTURE WORK

The efficiency of proposed methodology can be increased in several aspects like increasing efficiency, and computational complexity and security.

- Generating keys of larger size more than 128-bits and using AES encryption for handling this large key size.



- In this thesis, the proposed encryption technique based on symmetric keys cryptography. This work can be extended to asymmetric cryptography.
- Using asymmetric cryptography techniques, the authentication can also be done using digital signature.
- DNA hybridisation technique can be used in key generation phase, to ensure more security and randomness of the seed.(symmetric key)

## REFERENCES

---

- [1] M. Sugumaran, BalaMurugan. B, D. Kamalraj, "An Architecture for Data Security in Cloud Computing", World Congress on Computing and Communication Technologies,2014
- [2] <http://www.networkworld.com/article/2194263/tech-primers/authentication-in-the-cloud.html>
- [3] C. Wang, Q. Wang, K. Ren, N. Cao, and W. Lou, "Towards Secure and Dependable Storage Services in CloudComputing," IEEE Transactions on Services Computing, vol.5,no. 2, 2012,pp. 220-232.
- [4] T S Khatri and G B Jethava, "Survey on data Integrity Approaches used in the Cloud Computing, International Journal of Engineering Research & Technology, vol.1, Issue 9, November, 2012.
- [5] Saenger, Wolfram (1984). Principles of Nucleic Acid Structure. New York: Springer-Verlag
- [6] Watson JD, Crick FH (1953). "A Structure for Deoxyribose Nucleic Acid" (PDF). *Nature* 171 (4356): 737–738.
- [7] Alberts, Bruce; Johnson, Alexander; Lewis, Julian; Raff, Martin; Roberts, Keith; Walters, Peter (2002). *Molecular Biology of the Cell*; Fourth Edition. New York and London: Garland Science.
- [8] Clausen-Schaumann H, Rief M, Tolksdorf C, Gaub HE (2000). "Mechanical stability of single DNA molecules". *Biophys J* 78 (4): 1997–2007.
- [9] Crick, Francis (1988). "Chapter 8: The genetic code". *What mad pursuit: a personal view of scientific discovery*. New York: Basic Books. pp. 89–101
- [10] Kang Ning(2009),"A Psuedo DNA Cryptography Method", Cornell University Library
- [11] Debnath Bhattacharyya, Samir Kumar Bandyopadhyay," Hiding Secret Data in DNA Sequences", International Journal of Scientific & Engineering Research Volume 4(2013)
- [12] Mohammad Reza Abbasy, Bharanidharan Shanmugam," Enabling Data Hiding for Resource Sharing in Cloud Computing Environments Based on DNA Sequences",IEEE-2011

- [13] L. M. Pecora and T. L. Carroll, "Synchronization in Chaotic Systems", *Physical Review Letters*, vol. 64, no. 8, pp. 821–824, 1990
- [14] *Sync: The Emerging Science of Spontaneous Order*, Steven Strogatz, Hyperion, New York, 2003, pages 189-190.
- [15] [Lorenz, Edward N. (1963). "Deterministic non-periodic flow". *Journal of the Atmospheric Sciences* **20** (2): 130–141
- [16] G.Jakimoski and L. Kocarev, "Analysis of recently proposed chaos –based encryption algorithm", *Physics Letter*,A2001.
- [17] Wang, Xingyuan; Zhao, Jianfeng (2012). "An improved key agreement protocol based on chaos". *Commun. Nonlinear Sci. Numer. Simul.* 15 (12): 4052–4057.
- [18] Babaei, Majid (2013). "A novel text and image encryption method based on chaos theory and DNA computing". *Natural Computing. an International Journal* 12 (1): 101–107.
- [19] Wikipedia Chaos theory: Definition of chaos at Wiktionary;
- [20] Hasselblatt, Boris; Anatole Katok (2003). *A First Course in Dynamics: With a Panorama of Recent Developments*. Cambridge University Press
- [21] Alligood K.T,T.D Sauer,J.A Yorke, "Chaos an Intoduction to Dynamical systems",First ed 1996,New York:*Springer-Verlag*
- [22] Bertuglia C.S and F.Vaio,"Nonlinearity,Chaos & Complexity The Dynamics of Natural ans Social Systems",First ed.2005,United States:*Oxford University Press Inc*
- [23] Werndl, Charlotte (2009). "Are Deterministic Descriptions and Indeterministic Descriptions Observationally Equivalent?". *Studies in History and Philosophy of Modern Physics* 40 (3): 232–242.
- [24] M. S. Baptista, "Cryptography with Chaos",*Phys. Lett. A*, vol. 240, 1998.
- [25] R. Schmitz and J. Franklin, "Use of Chaotic Dynamical Systems in Cryptography", vol. 338, 2001
- [26] Kotulski Z, Szczepariski J. Discrete chaotic cryptography (DCC). In: Proc NEEDS 97
- [27] Shiguo Lian, Jinsheng Sun, Zhiquan Wang(), "A block cipher based on a suitable use of the chaotic standard map", *Science direct* (2004)
- [28] FIPS PUB 180-1, "Secure Hash Standard", Federal Information Processing Standard(FIPS), Publication 180 -1, National Institute of Standards and Technology, US Department of Commerce, Washington D.C., 1995.

- [29] Lu HP, Wang SH, Hu G. Pseudo-random number generator based on coupled map lattices. *Int J Modern Phys B* 2004;18(17–19): 2409–14
- [30] Sh. Li, X. Mou and Y. Cai, “Pseudo-Random Bit Generator Based on Couple Chaotic Systems and its Applications in Stream-Cipher Cryptography”, *INDOCRYPT 2001*, LNCS, Springer-Verlag, Berlin, 2001
- [31] Whitfield Diffie, Martin E Hellman, ”New Directions In Cryptography”, *IEEE Transactions On Information Theory*, Vol.It-22,No.6,November 1976.
- [32] H. J. Shiu, K. L. Ng, J. F. Fang, R. C. T. Lee and C. H. Huang, “Data hiding methods based upon DNA sequences”, *Information of Science*, vol.180, no.11, pp.2196-2208, 2010.
- [33] D.Erdmann and S.murphy, ”HENON STREAM CIPHER”,*Electronics Letters*,23<sup>rd</sup> april 1992 vol.28 no.9
- [34] Zeng X.,R.A Pielke and R. Eykholt, “Chaos theory and its application to the Atmosphere”,*Bulletin of the American Meteorological Society*,1993.74(4):p.631-639
- [35] Claude E. Shannon, "Communication Theory of Secrecy Systems", *Bell System Technical Journal*, vol. 28-4, page 656–715, 1949.
- [36] Wikipedia source, ”[https://wiki/Confusion\\_and\\_diffusion](https://wiki/Confusion_and_diffusion)”,
- [37] Qiang Zhang \*, Ling Guo, Xianglian Xue, Xiaopeng Wei,"An Image Encryption Algorithm Based on DNA Sequence Addition Operation",*Key Laboratory of Advanced Design and Intelligent Computing* ,2011
- [38] Jin-Shiuh Taur<sup>1</sup>, Heng-Yi Lin<sup>1</sup>, Hsin-Lun Lee<sup>1</sup> and Chin-Wang Tao, “Data Hiding in DNA Sequences Based On Table Lookup Substitution”, *International Journal of Innovative Computing, Information and Control*, Volume 8, Number 10(A), October 2012
- [39] K. Menaka, "Message Encryption Using DNA Sequences", *Department of IT and Applications, World Congress on Computing and Communication Technologies*,2014
- [40] Salah H. Abbdal, Hai Jin, Deqing Zou, Ali A. Yassin,” Secure and Efficient Data Integrity Based on Iris Features in Cloud Computing”,*7th International Conference on Security Technology*,2014

- [41] Hyotaek Lim<sup>2</sup>, Hoon Jae-Lee<sup>2</sup>, Amlan Jyoti Choudhury, Pardeep Kumar, Mangal Sain, "A Strong User Authentication Framework for Cloud Computing", IEEE Asia -Pacific Services Computing Conference, 2011
- [42] G.SudhaSadasivam, K.AnithaKumari,S.Rubika,"A Novel Authentication Service for Hadoop in Cloud Environment",IEEE Transaction on Information Security,vol no 20,pp 329-340,2012
- [43] M. L. Das, A. Saxena, and V. P. Gulati, "A dynamic ID-based remote user authentication scheme" IEEE Transactions on Consumer Electronics, vol. 50, no. 2, pp. 629-631, 2004.
- [44] Jen-Ho Yang, Pei-Yu Lin, "An ID-Based User Authentication Scheme for Cloud Computing", Tenth International Conference on Intelligent Information Hiding and Multimedia Signal Processing, 2014
- [45] Salah H. Abbdal, Hai Jin, Deqing Zou, Ali A. Yassin, "Secure and Efficient Data Integrity Based on Iris Features in Cloud Computing", 7th International Conference on Security Technology, 2014
- [46] Faraz Fatemi Moghaddam, Touraj Khodadadi, Rama Roshan Ravan, "SUAS: Scalable User Authentication Scheme for Secure Accessing to Cloud-Based Environments", IEEE Symposium on Computer Applications & Industrial Electronics (ISCAIE) , 2014

