A Multilevel Security Scheme based on UNICODE and RGB Color Model using DNA Cryptography

Panchami V.
Asst. Professor
Computer Science & Engineering
Toc H Institute of Science & Technology, Arakkunam,
Ernakulam, India

Fasila K.A.
Lecturer
Computer Science & Engineering
Toc H Institute of Science & Technology, Arakkunnam,
Ernakulam, India

Sudhin Vamattam
Team Leader
GSC, Ernst & Young Infopark,
Cochin,
Kerala, India

ABSTRACT

RGB color oriented DNA based computing is a new research area in cryptography. In this paper a formal approach to encrypt and decrypt information using colors is proposed. An information or message consists of set of Characters, Symbols, and Digits. Encoding is done using UNICODE, so that any language can be encrypted. After encoding, the message is encrypted and compressed using RGB colors. Each color has its own hexadecimal values; these hexadecimal values are converted into binary values. These binary values are treated as 2 parts- message and the key. Then the message and the key are compressed by performing XOR operation and the key is encrypted using RSA algorithm. Finally the message is encoded as DNA codons and then transmitted to the receiver.

DNA can be used not only to store and transmit information, but also to perform computations. The fundamental idea behind this encryption technique is to enforce DNA concept in cryptographic algorithms and to open the door for applying the DNA and Amino Acids concepts to other conventional cryptographic algorithms so that we can ensure multi-layer security.

General Terms

Security, DNA Cryptography, Encryption, Decryption

Keywords

DNA Cryptography, UNICODE Encoding, RGB color model, Compression, Basic principles of Central dogma of molecular biology.

1. INTRODUCTION

The design of a strong encryption mechanism which is suitable for Multilanguage messages is a tedious task. The process of developing such an algorithm involves complex problems in order to ensure the security of data for at least a certain period of time. UNICODE is a computing industry standard for the consistent representation and handling of text expressed in most of the world's writing systems. Developed in conjunction with the Universal character Set standard and published in book form as The UNICODE Standard, the latest version of UNICODE consists of a repertoire of more than 107,000 characters covering 90 scripts, a set of code charts for visual reference, an encoding methodology and set of standard character encodings, an enumeration of character properties such as upper and lower case, a set of reference data computer files, and a number of related items. This paper introduces a new technique for cryptography by using the concept of DNA computing in collaboration with UNICODE and colors in universe. We can see about 1000 levels of light and dark, 100 levels of red and green, and 100 levels of yellow and blue for a single viewing condition in a laboratory. This means that the total number of colors we can see is 10 million colors [6]. A computer can display about 16.8 million colors to create full color pictures, really more than necessary for most situations. The appearance of a color is greatly affected by the viewing conditions, which include the color of the lighting, the amount of lighting, and other colors in the scene. Colors also appear in different modes when they appear on different objects such as surfaces, light sources, or within volumes [7]. Since we can see at least 10-million colors in a single viewing condition and the variety of viewing conditions and observers is endless, then the only truly correct answer is infinity. If we have 10-million colors, times 10-million lighting types, times 10-million lighting levels, 10-million surrounding colors, times 6-billion people in the world, using these modes of viewing we get a really huge number. Each color is unique and can be represented in hexadecimal values, so we can represent any data or user in colors, this is the advantage that we see while using colors in our encryption and decryption algorithm. The best way of achieving a robust system is to act on scalability that is to reach a large scale complexity for the problem. Such task can be made possible and handled by the innovative DNA computing, which allows a very high degree of parallelism on one hand and a huge storage capacity on the other hand. DNA is a nucleic acid that contains the genetic instructions used in the development and functioning of all living organisms and some viruses. DNA computing was born with the Adelman's pioneering work [2].

The vast parallelism and the density of information inherent to DNA are exploited to solve hard problems in different axis in computer science. In the field of cryptography, Gehanietal [1] introduces the first algorithm of DNA based cryptography, followed by many others [3] [4] [5]. In this work, we are not intended to use real DNA strands but we took the DNA concept. DNA-based algorithm called YAEA was proposed by Aminet al [1] in which a binary form of data, such as plaintext messages, and images are transformed into sequences of DNA nucleotides. DNA nucleotides that represent the binary octet plain text character within a Canis Familiaris genomic chromosome. In this paper we are not dealing with the real DNA's but using the concept only. This paper is organized as follows: Section 2 describes the proposed algorithm for encryption. Section 3 illustrates the encryption method with example. Decryption mechanism is explained in section 4 and section 5 deals with conclusion and future enhancements.

2. PROPOSED ALGORITHM

The proposed algorithm proceeds through mainly 7 steps:

- Encoding of plain text using UNICODE.
- Encrypt using RGB color model.
- Compression of colors.
- Conversion to binary equivalent form.
- Separation of message & key and the key encryption using RSA[6] algorithm
- Final encryption done by the concept based on DNA cryptography.
- Convert the cipher text obtained to colors

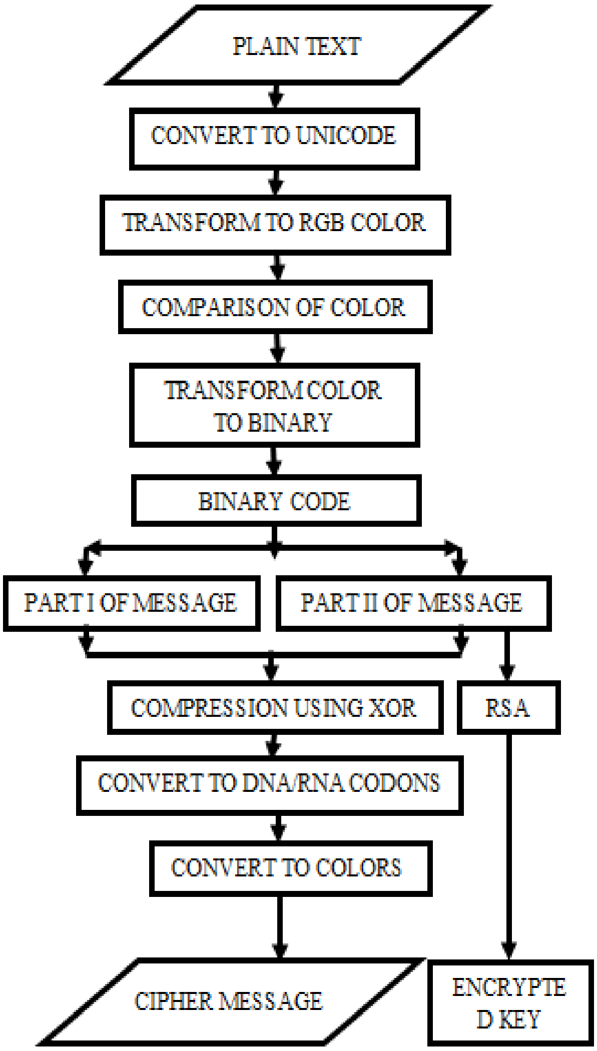


Figure 1: Flow chart of the proposed algorithm

3. ENCRYPTION

3.1 Encode Plaintext using UNICODE

The characters, symbols, digits of any language are encoded using UNICODE [7]. The objective of UNICODE is to unify all the different encoding schemes so that confusion between computers can be limited as much as possible. The advantage of UNICODE compared to ASCII is that it can encode characters from multilingual plane. ASCII can encode characters only from English language. These days the UNICODE standard defines values for over one million characters and can be seen at the UNICODE Consortium. It has several character encoding forms, UTF standing for UNICODE Transformation Unit.

UTF-8 only uses one byte (8 bits) to encode English characters. It can use a sequence of bytes to encode the other characters. UTF-8 is widely used in email systems and on the Internet. UTF-16 uses two bytes (16 bits) to encode the most commonly used characters. If needed, the additional characters can be represented by a pair of 16-bit numbers. UTF-32 uses four bytes (32 bits) to encode the characters. It became apparent that as the UNICODE standard grew a 16-bit number is too small to represent all the characters. UTF-32 is capable of representing every UNICODE character as one number. In this proposed algorithm, we are using UTF-16 for encoding the plain text to UNICODE values.

3.1.1 Code Points

A code point is the value that a character is given in the UNICODE standard. The values according to UNICODE are written as hexadecimal numbers and have a prefix "U+". For example, to encode the characters I looked at earlier, "A" is U+0041, "1" is U+0031, "#" is U+0023. These code points are split into 17 different sections called planes. Each plane holds 65,536 code points. The first plane, which holds the most commonly used characters, is known as the basic multilingual plane.

3.1.2 Code Units

The encoding schemes are made up of code units. They are way to provide an index for where a character is positioned on a plane. For instance, with UTF-16 each 16-bit number is a code unit. The code units can be transformed into code points. For example, the flat note symbol " " has a code point of U+1D160 and it lives on the second plane of the UNICODE standard. It would be encoded using the combination of the following two 16-bit code units: U+D834 and U+DD60.

In this proposed new policy, first of all it checks each and every character in the given file. Then it finds what equivalent UNIODE of each character is.

3.2 Encrypt Using RGB Color Model

The characters in the plain text are translated into colors. After encoding, each and every characters, symbols, digits of any language are encrypted into colored-charts. The UNICODE of each character is converted into colors using following Table 1, which provides a mapping of UNICODE and colors. A computer displays about 16.8 million colors to create full color pictures, really more than necessary for most situations. Now we are considering 10 million Colors only. And the UNICODE standard defines values for over 100,000 characters and can be seen at the UNICODE Consortium. Now 10 million colors and 100,000 characters are available in a computer system [6]. Now we can create a dynamic mapping table. As we said earlier, computer screen supports millions of colors, so we can create millions of dynamic color-chart tables. In each table there are four columns unique id, the data, the Unicode of that data and the color. Each table is identified by unique id number, that unique id number acts as first private key (KEY 1) while decrypting the message. This provides more difficulty in decryption side. According to the Table 1, an example is as follows:

Table 1: Colored Chart Table

Unique ID for each color	Character (any language) / Symbols /digits	UNICODE	Colors
1	A	U+0041	🟥
2	B	U+0042	🟦
3	C	U+0043	🟫
4	D	U+0044	🟩
5	E	U+0045	🟨
. .	. .	. .	. .
25	1	U+0031	🟩
26	2	U+0032	🟦
. .	. .	. .	. .
33	#	U+0023	⬛
34	Space	U+0020	🟥

Consider the character "A". The UNICODE of "A" is U+0041.This UNICODE is converted into the equivalent color which is black. Similarly "B" is U+0042 which is blue, "D" is U+0044 which is green, "E" is U+0045 which is yellow. The following Figure 2 shows color code equivalent of the word 'BEAD'.

After RGB color Encryption:

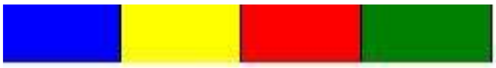


Figure 2: Color code equivalent to the word BEAD

We will also check if the two adjacent colors is same or not, that means if the plain text have adjacent same characters, then we will replace it with another color as in the given table, (Table 2). For each time the adjacent color arrives it is replaced with each unique color as in Table 2, the time is taken as second private key (KEY 2).The aim of this step is to achieve compression, and can reduce the size of the encrypted data. This compression technique is very simple as well as its very efficient and effective for any type of data.

Table 2: Colored chart table

No. of times same color appears continuously	Color
1	🟪
. .	
n	🟪

3.3 Compression of colors

The number of colors in the resulting set after the first step will be same as the number of characters in the plain text. But for a huge amount of data we will get a combination of a lot of colors for which the maintenance task is complicated one and not efficient. So we go for compression by mixing adjacent two colors. Each color has its own unique hexadecimal value [7]. If we mix two colors we will get another color which has a unique hexadecimal value, which means the resultant color can be produced only by mixing with those two colors. Each color has its own unique ID. This unique ID is used at the time of decryption to retrieve the component colors [6]. This set of unique IDs is treated as the third key, KEY3. So we input the hexadecimal value of the resultant color to a color blending tool, it will give those two colors with their hexadecimal values. Generation of keys: KEY 1, KEY 2, KEY 3. The three keys are sent to the receiver using key agreement proposed by Diffie-Hellman [2]. From the above example, mixing of two adjacent colors is shown in figure 3.

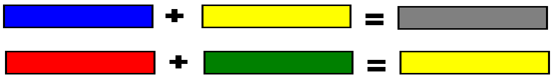


Figure 3: Compression of colors

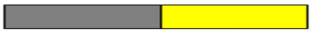
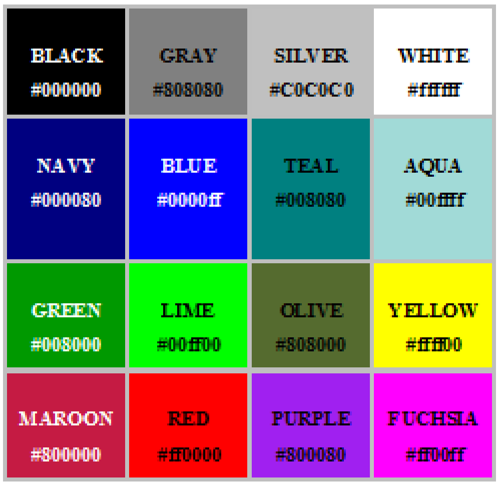


Figure 4: Result after 1st stage of compression

3.4 Conversion of colors to binary form

Every color in RGB model has an equivalent hexadecimal code [7]. Color codes are hexadecimal triplets representing the colors red, green and blue (#RRGGBB). For example, for red color, code is #FF0000, which is '255' red, '0' green and '0' blue. See Figure 5.



BLACK #000000	GRAY #808080	SILVER #C0C0C0	WHITE #ffffff
NAVY #000080	BLUE #0000ff	TEAL #008080	AQUA #00ffff
GREEN #008000	LIME #00ff00	OLIVE #808000	YELLOW #ffff00
MAROON #800000	RED #ff0000	PURPLE #800080	FUCHSIA #ff00ff

Figure 5: Colors and their hexadecimal values

For further proceeding, we need to convert this hexadecimal code to a binary equivalent form. Now the hexadecimal code of the final color is considered and this code is converted to the binary form through simple hex to binary conversion. Now, the binary value we obtained is divided and treated as two parts-1.The message 2.The key

3.5 Compression of data

The message and key that we obtained in binary format is again compressed by performing simple XOR operation on bits [8]. So that we can reduce the size of encrypted message .Since the compression using XOR is being performed, we can easily extract the message and key later by using reverse XOR operation. The key is now encrypted using RSA [9] algorithm.

3.6 Encryption using DNA cryptography

The encryption process starts by the binary form of data (message or image) which is transferred to DNA form according to Fig 4. Then the DNA form is transferred to the mRNA. This mRNA is again converted to amino acids according to Table 3 which is a standard universal table of Amino acids and their codons presentation in the form of DNA-RNA codon table.

Table 3: DNA representation of binary values

Binary Value	DNA Digital Coding
00	A
01	C
10	G
11	T

In the information science, the most fundamental coding method is binary digital coding. This means any data can be encoded as a combination of 0s and 1s. There are 4 kinds of bases. They are ADENINE (A), THYMINE (T), CYTOCINE (C) and GUANINE (G) in DNA sequence [10]. The simplest coding pattern to encode nucleotide bases- A, T, G, c is by means of 4 digits-0 (00), 1(01), 2(10), 3(11). There are 24 possible patterns. Instead of theses binary numbers, we use encoding using A, C, G and T.

The genetic code can be expressed as either RNA codons or DNA codons. RNA codons occur in messenger RNA (mRNA) and are the codons that are actually "read" during the synthesis of polypeptides (the process called translation)[9]. But each mRNA molecule acquires its sequence of nucleotides by transcription from the corresponding gene. The DNA Codons is read the same as the RNA codons Except that the nucleotide thymidine (T) is found in place of uridine (U) [10]. So in DNA codons we have (TCAG) and in RNA codons, we have (UCTG).A gene is a sequence of DNA that contains genetic information and can influence the phenotype of an organism. Within a gene, the sequence of bases along a DNA strand defines a messenger RNA sequence, which then defines one or more protein sequences [11]. The relationship between the nucleotide sequences of genes and the amino-acid sequences of proteins is determined by the rules of translation, known collectively as the genetic code [12]. The genetic code consists of three-letter 'words' called codons formed from a sequence of three nucleotides (e.g. ACT, CAG, TTT).

In transcription, the codons of a gene are copied into messenger RNA by RNA polymerase. This RNA copy is then decoded by a ribosome that reads the RNA sequence by base-pairing the messenger RNA to transfer RNA, which carries amino acids. Since there are 4 bases in 3- letter combinations, there are 64 possible codons (4^3 combinations). These encode the twenty standard amino acids, giving most amino acids more than one possible codon. There are also three 'stop' or 'nonsense' codons signifying the end of the coding region; these are the TAA, TGA and TAG codons.

Our proposed DNA encryption is based on the figure 7. This is a dynamic table. We have randomly distributed the 64 codons in various places of a particular message to encrypt. To encode another message we will use the same table but with another random distribution of codons. For each distribution, we have a unique ID so that while decrypting the code, the recipient will get the id along with KEY1.

3.7 Conversion of cipher text to cipher color message

As the final step, the cipher text is encoded as colors by referring to the Figure 1.

4. ILLUSTRATION WITH EXAMPLE

In this example the Plain Text is BEAD.

4.1 UNICODE of each Character

B ➔U+0042

E ➔U+0045

A ➔U+0041

D ➔U+0044

4.2 Conversion to RGB colors

According to the Fig 2, the UNICODE of each character in the 'BEAD' is encoded to RGB colors as follows:

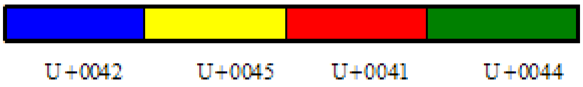


Figure 6: UNICODE is encoded to RGB Colors

4.3 Compression of colors

Adjacent two colors are mixed to achieve compression. Blue and yellow is mixed to get ash color, red and green is mixed to get yellow.

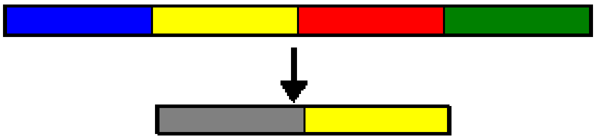


Figure 7: Result after the compression

4.4 Hexadecimal to binary conversion

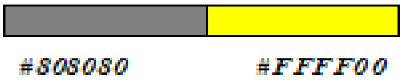


Figure 8: Result after the compression

These hexadecimal values of each color are converted into binary values.

#808080 : 100000001000000010000000

#FFFF00 : 111111111111111100000000

Hence, the binary equivalent form of our plain text

'BEAD' is:

100000001000000010000000111111111111111100000000

This binary value is then compressed.

4.5 Compression of Data

The binary form of the plain text is now divided into 2 parts. They are the message and the key

Message : 100000001000000010000000

Key : FFFF00

The key value obtained is encrypted using RSA [9] algorithm and sent to the recipient. The compressed form using XOR is given below:

011111110111111110000000

This binary value act as the input is encrypted using DNA concept.

4.6 DNA Based Encryption

4.6.1 Conversion from binary to DNA form

The compressed binary form is now converted into DNA representation using DNA digital coding as shown in Table 3.

011111110111111110000000

↓

CTTTCTTTGAAA

4.6.2 DNA based Transcription

This set of DNA codons are transcript to the following form:

CTT TCT TTG AAA

4.6.3 DNA based Translation

These DNA codons are translated into mRNA codons.

CUU UCU UUG AAA

4.6.4 Final Conversion to Cipher Text

By referring the table shown in Table 4, we got the encrypted form of the plain text. Final cipher text is 'LSFF'.

In the next step this cipher text is again encoded to cipher colors.

4.7 Creation of cipher colors

According to the Figure 2, the UNICODE of each character in the 'LSFF' is encoded to cipher colors .This step is optional because the cipher text along gives much security, we can avoid the color encoding technique is used in our algorithm from the opponent.

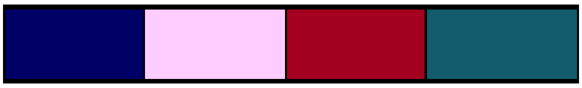


Figure 9: The Cipher Colors

5. DECRYPTION

The decryption process is simply the inverse of the encryption process. Here, we use a key set which consists of three keys. They are KEY1, KEY2 and KEY3. These keys are exchanged using Diffie-Helmann [2] key exchange algorithm. In addition to the key set, we use another key which is encrypted by RSA [9] algorithm.

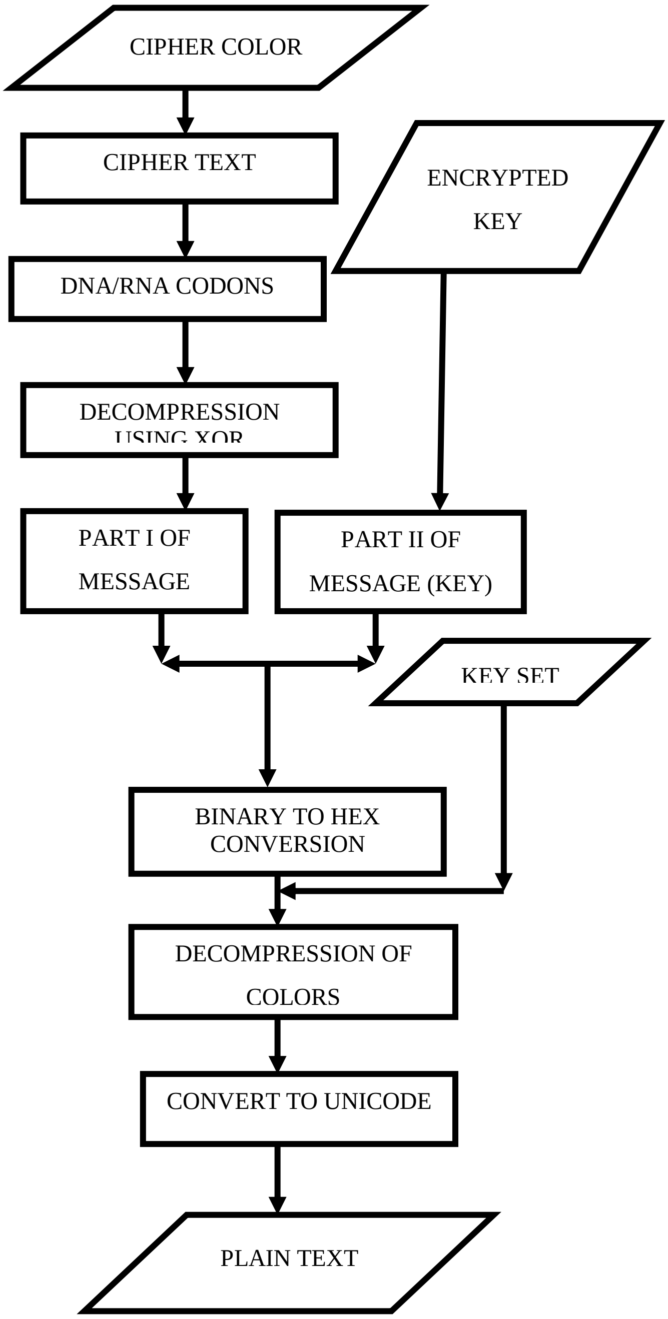


Figure 10: Flow chart of the proposed decryption algorithm

Table 4: DNA Encoding Table

K	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	W	X	Y	Z
1	C G U	A G A	G C U	G C A	G G U	C A A	A A U	A U G	A U U	C A U	U G G	G A U	G A A	A C U	A U G	C C U	U G G	A A A	U U A	U A A	U C U	U A C	U C C	U U U	C U A
2	C G C	A G G	G C C	G C C	G G C	A A C	A A C		A U C	C A C	U G C	G A C	G A G	A C C		C C C		A A G	U U G	U G A	U C A	G U C	U C G	U U C	C U G
3	C G A				G G A				A U A					A C A		C C A			C U U	U A G	A G U	G U G	A G C		
4	C G G				G G G									A C G		C C G			C U C						

6. CONCLUSION AND FUTURE WORK

The proposed algorithm offers multilayer security. This technique is to open the door for the idea of applying the concepts of UNICODE, DNA and Amino Acids and colors to other conventional cryptographic algorithms to enhance their security features. We are using UNICODE so that we encrypt thousands of characters, symbols, digits etc of any language instead of the ASCII code. Then we are encoding with RGB colors. The advantage is that the display screen supports millions and millions of colors. Each color is unique. In the proposed algorithm, we are mixing the adjacent two colors. We can reduce the length to the half of the length of original message. If we are able to combine a group of colors and form a single color, we can represent messages of any length using just one color. So we can compress huge amount of data. In future we are planning to implement this encoding scheme on other known algorithms and measuring its performance and security. Experiments should be conducted to implement the algorithm on different applications to ensure its feasibility and applicability.

7. REFERENCES

[1] Kang Ning, "A Pseudo DNA Cryptography Method", Independent Research Study Project for CS5231, October 2004.

[2] Leonard Adleman, "Molecular Computation of Solutions to Combinatorial problems", Science 266: 1021-1024, November 1994.

[3] TAYLOR Clelland Catherine, Vivjana Risca, Carter Bancroft, 1999, "Hiding Messages in DNA microdots", Nature Magazine, Vol 399, June10, 1999.

[4] Sherif T.Amin, Magdv Saeb, Salah EI-Gindi, "A DNA-based Implementation of YAEA Encryption Algorithm", International Conference on Computational Intelligence (CI2006), San Franciso, Nov20, 2006.

[5] Dominik Heider and Angelika Barnekow, "DNA- based watermarks using the DNA-Crypt algorithm", Published: 29 May 2007 BMC.

[6] Piyush Marwaha and Parsh Marwaha, "Visual cryptographic steganography in images", Second International Conference on computing, Communication and Networking technologies, 2010.

[7] S.Pavithra Deepa, S.Kannimuthu, V.Keerthika, "Security using colors and Armstrong numbers", Innovations in Emerging Technology (NCOIET), Feb.2011.

[8] Ashish CreForman, Thomas LaBean and John Reif, "DNA-Based Cryptography", DIMACSDNA Based Computers V, American Mathematical Society, 2000.

[9] Francis crick, "Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid." April 25, 1953.

[10] Bruce Alberts, Alexander Johnson, Julian Lewis, Martin Raff, Keith Roberts, Peter Walter, Molecular Biology of the Cell, Fourth Edition Garland Publishing, 2002.

[11] Souhila Sadeg "An Encryption algorithm inspired from DNA" IEEE pp 344 – 349 November 2010.

[12] Hayam Mousa et al. , Data Hiding Based on Contrast Mapping Using DNA Medium, The International Arab Journal of Information Technology, Vol. 8 No. 2 , 2011.