Secure Data Transfer through DNA Cryptography using Symmetric Algorithm

Bonny B. Raj
Research Scholar
Department of Computer Science and Engineering
Hindustan University
Chennai, India

J. Frank Vijay, PhD
HoD in Information Technology
KCG College of Engineering and Technology
Chennai, India

T. Mahalakshmi, PhD
Principal
Sree Narayana Institute of Technology, Vadakkevila,
Kollam, India

ABSTRACT
This paper presents a novel symmetric algorithm in the area of DNA cryptography. Secure Data Transfer is an important factor for data transmission. The transmission of information can be of local or of global scope. But it is mandatory to secure information from unauthorized access. Security is very important factor encryption. This method proposes a secured symmetric key generation process which generates initial cipher and this initial cipher is then converted into final cipher using random key generated DNA sequences, so as to make it complicated.

General Terms
Bio- inspired computing, Security, Algorithms et. al.

Keywords
Bio-inspired computing, Symmetric Encryption, Random key, Symmetric key.

1. INTRODUCTION
Security of valuable information is ensured using cryptographic techniques. Using such techniques enables the sender for secure transmission as well as storage of sensitive information through the internet. A cryptographic system applies encryption of information and thus produces an encrypted output which may be meaningless to an intruder who has no knowledge of the key. The knowledge of the key is essential factor for encryption and decryption process. The fundamental tool for cryptography is a simple function which is one-way and is easy to compute but very hard to get invert [1].

Encryption and decryption phases of cryptography are determined by keys. Based on keys, cryptographic systems can be classified as Symmetric Key Cryptography (SKC) and Asymmetric Key Cryptography (AKC) also known as public key cryptography [2]. Symmetric Key Cryptographic system uses single key for both encryption and decryption. In Asymmetric Key Cryptographic system different keys are used for both encryption and decryption.

SKC is based on sharing secrecy of the key, where as AKC is based on personal secrecy of the key [3].Generally in the former symbols are permuted or substituted where as in the later numbers are manipulated using functions. In the SKC key is shared between the sender and the receiver. In AKC the key is personal and is known to themselves. The difficulty in SKC approach is the distribution of the key where as in AKC there is no such distribution of the key [12].

DNA Cryptography is generally defined as hiding data in DNA sequence. The advantages of this is as follows [9].

a) By supporting large parallelism which helps to improve computational speed.

b) The molecules of DNA acts as carrier of transmission with large capacity.

c) Power consumption is small.

In the present work a novel SKC algorithm using DNA cryptography is proposed which does not need any DNA Chromosome OTP structure as seen in paper [12].

The remaining part of this paper is organized as follows: Section 2 gives background information relevant to the proposed work which is followed by literature review in section 3.In section 4 a proposed algorithm is elucidated followed by discussion in section 5 and conclusion in section 6.

2. BACKGROUND
This section describes some terms which are frequently used in this paper.

2.1 DNA (De-oxyribo Nucleic acid)
DNA refers to De-oxyribo nucleic acid is a nucleic acid which contains genetic information. It is used for the growth as well as functioning of all living organisms. They have two long chains of nucleotide which has a double helical structure [5].It a collection of most complex organic molecules. The instructions in DNA is required for the construct other components of cells such as protein, RNA molecules etc [16]. The segments in DNA that hold genetic information are known as genes. These sequences are used modifying the use of this genetic information, just as a string of binary information which is encoded with zeroes and ones. DNA strand encoding include four bases which are represented by letters A (Adenine), T (Thymine), C (Cytosine) and G (Guanine) [15].Each alphabet is related to a nucleotide. They are very long. For instance, the DNA sequence of length 10 nucleotides long can be represented as ATCGAATTCG.

2.2 DNA Cryptography
DNA cryptography is a new promising direction in cryptography research that emerged with the evolution of DNA computing. DNA can be used for storing, transmitting and computation of information [6]. Parallelism and extraordinary information density inbuilt in this molecule are exploited for encryption and decryption purpose[14]. Different DNA based algorithm are available for this purpose. In this paper, the research conducted by various authors related to this discipline is taken into consideration. This shows, how DNA cryptography uses DNA as a

computational tool to manipulate for encryption and decryption[7]. Research work can be done on DNA by two ways either by molecule or by simulation [8].

2.3 Symmetric key Algorithm

Symmetric or Secret key can be classified as stream ciphers or block ciphers. Stream ciphers operate on a single bit (byte or computer word) at a time, and implement some form of feedback mechanism so that the key is constantly changing [R1].A block cipher encrypts a block of data at a time using the same key on each block. There are a number of other secret-key cryptography algorithms that are also in use today like CAST-128 (block cipher), RC2 (block cipher) RC4 (stream cipher), RC5 (block cipher), Blowfish (block cipher), Two fish (block cipher). In 1997, NIST initiated a process to develop a new secure cryptosystem for U.S. government applications. The proposed system use symmetric key stream cipher algorithm.

3. LITERATURE REVIEW

The literature study revealed the existence of the paper titled" "Yet Another Encryption Algorithm" (YAEA) developed by Saeb and Baith [13] which is based on a conventional symmetric key algorithm. The study also reveal the existence of another paper based on DNA computing in the field of cryptography[12].

It was found that the field of DNA Cryptography belongs to the area of Bio-inspired computing methodologies. Biologically-inspired computing (BIS) is the study which deals with the topics of connection, social behavior and its emergence [4]. Bio-inspired computing methodologies are designed to solve specific problem, biological system, which follows a specific procedure and can has similar properties. Some forms shows complex behavior and rules. Complexity increases until the result is something complex. Final results are often completely different from what the original rules would be expected to produce.

The first DNA based cryptographic technology was seen in [6].Somewhat similar to the proposed work is seen in [13].

Cryptosystems which use a secret random OTP are known to be perfectly secure [12]. OTP encryption uses a large non-repeating set of truly random key letters. Each pad is used exactly once, for only one message.

The survey also revealed the different DNA data encryption algorithm which uses traditional mathematical operations and/or data manipulating DNA techniques [10].The proposed method is also an application in this area.

4. PROPOSED SYSTEM

Proposed system is a new encryption algorithm based on random key generation of DNA pattern. There are three stages in this algorithm- Encryption, Random Key Generation and Decryption.

In the first stage the source data is encrypted which is the input to the second stage. In second stage random key is generated, say Pk which is used for next level of encryption. In the third stage decryption process takes place.

Fig: 1 gives the flow chart of the stages I and II of the proposed system and the Fig: 2 gives the flow chart of the stage III of the proposed system.

The remaining part of this section explains in details each of these three stages.

4.1 Encryption

This is the stage one of the proposed algorithm. The input for this is the source data which is in the text format. Each character of the source data is converted to its corresponding ASCII value which is in turn converted to its binary value in Fig 1 first three blocks corresponds to this stage. This process is explained in details as follows.

Consider the message the M to be transmitted to the receiver , the different steps for encryption is as follows

Step 1: The text to be transmitted is first converted to ASCII code (in decimal format).

Step 2: These decimal values are grouped into blocks.

Step 3: This encoded ASCII message (decimal values) is then converted to binary format (0's and 1's).

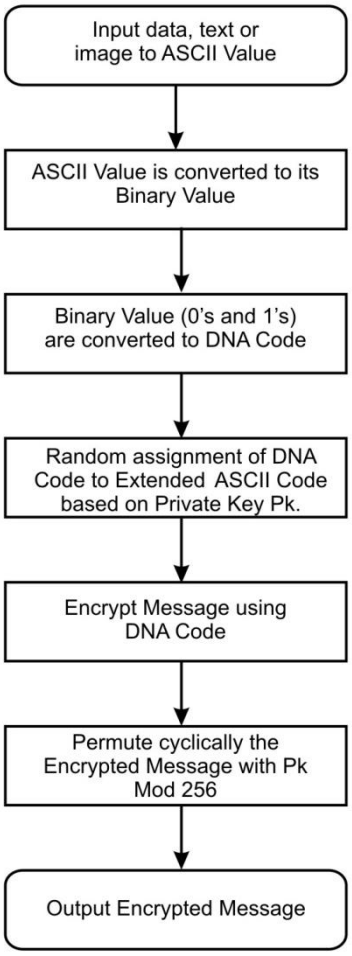

Fig1: Stage I and II of the Proposed System.

4.2 Random Key Generation

This is the second stage in the proposed system. In this stage a random key say Pk in the range 1 to 256 is generated for further encryption. Corresponding to each value of Pk an index table is of size 256 is generated where each value of the table corresponds to a combination of A, T, G and C .When Pk =1 the index table generated is as shown in Table 1. The size of the index table is 256 which corresponds to the permutation of four characters. A, T, G and C. As the value of Pk varies the index table will also varies.

The encrypted binary data from the stage 1 is the input for this stage. First the input data is selected as pairs. Each pairs is replaced by the DNA nucleotides A, T, G, and C corresponding to the values 00,01,10 and 11.

From the index table the index key corresponding to the combination of A, T, G, and C is obtained which is the final encrypted value corresponding to the character of source data.

In Fig:1 blocks 4,5,6,7 represents this stage. The value of Pk is also transmitted along with encrypted character from this source data. Given below in detail the various steps involved in this stage.

Step 4: Encrypted binary data obtained from the first stage is considered in the form of pairs. Since the encrypted data is binary these pairs will be either 00 or 01 or 10 or 11.These pairs are substituted by A for 00, T for 01, G for 10, and C for 11.

Step 5: Generate Pk

Step 6: Corresponding to the value of Pk generate index table.

Step 7: The A, T , G and C obtained from step 4 is used to find the final encrypted key from the index table generated in step 6.

Table 1: Random Generated Key

1	AAAA	33	CAAA	65	GAAA	97	TAAA	129	AGAA	161	CGAA	193	GGAA	225	TGAA
2	AAAC	34	CAAC	66	GAAC	98	TAAC	130	AGAC	162	CGAC	194	GGAC	226	TGAC
3	AAAG	35	CAAG	67	GAAG	99	TAAG	131	AGAG	163	CGAG	195	GGAG	227	TGAG
4	AAAT	36	CAAT	68	GAAT	100	TAAT	132	AGAT	164	CGAT	196	GGAT	228	TGAT
5	AACA	37	CACA	69	GACA	101	TACA	133	AGCA	165	CGCA	197	GGCA	229	TGCA
6	AACC	38	CACC	70	GACC	102	TACC	134	AGCC	166	CGCC	198	GGCC	230	TGCC
7	AACG	39	CACG	71	GACG	103	TACG	135	AGCG	167	CGCG	199	GGCG	231	TGCG
8	AACT	40	CACT	72	GACT	104	TACT	136	AGCT	168	CGCT	200	GGCT	232	TGCT
9	AAGA	41	CAGA	73	GAGA	105	TAGA	137	AGGA	169	CGGA	201	GGGA	233	TGGA
10	AAGC	42	CAGC	74	GAGC	106	TAGC	138	AGGC	170	CGGC	202	GGGC	234	TGGC
11	AAGG	43	CAGG	75	GAGG	107	TAGG	139	AGGG	171	CGGG	203	GGGG	235	TGGG
12	AAGT	44	CAGT	76	GAGT	108	TAGT	140	AGGT	172	CGGT	204	GGGT	236	TGGT
13	AATA	45	CATA	77	GATA	109	TATA	141	AGTA	173	CGTA	205	GGTA	237	TGTA
14	AATC	46	CATC	78	GATC	110	TATC	142	AGTC	174	CGTC	206	GGTC	238	TGTC
15	AATG	47	CATG	79	GATG	111	TATG	143	AGTG	175	CGTG	207	GGTG	239	TGTG
16	AATT	48	CATT	80	GATT	112	TATT	144	AGTT	176	CGTT	208	GGTT	240	TGTT
17	ACAA	49	CCAA	81	GCAA	113	TCAA	145	ATAA	177	CTAA	209	GTAA	241	TTAA
18	ACAC	50	CCAC	82	GCAC	114	TCAC	146	ATAC	178	CTAC	210	GTAC	242	TTAC
19	ACAG	51	CCAG	83	GCAG	115	TCAG	147	ATAG	179	CTAG	211	GTAG	243	TTAG
20	ACAT	52	CCAT	84	GCAT	116	TCAT	148	ATAT	180	CTAT	212	GTAT	244	TTAT
21	ACCA	53	CCCA	85	GCCA	117	TCCA	149	ATCA	181	CTCA	213	GTCA	245	TTCA
22	ACCC	54	CCCC	86	GCCC	118	TCCC	150	ATCC	182	CTCC	214	GTCC	246	TTCC
23	ACCG	55	CCCG	87	GCCG	119	TCCG	151	ATCG	183	CTCG	215	GTCG	247	TTCG
24	ACCT	56	CCCT	88	GCCT	120	TCCT	152	ATCT	184	CTCT	216	GTCT	248	TTCT
25	ACGA	57	CCGA	89	GCGA	121	TCGA	153	ATGA	185	CTGA	217	GTGA	249	TTGA
26	ACGC	58	CCGC	90	GCGC	122	TCGC	154	ATGC	186	CTGC	218	GTGC	250	TTGC
27	ACGG	59	CCGG	91	GCGG	123	TCGG	155	ATGG	187	CTGG	219	GTGG	251	TTGG
28	ACGT	60	CCGT	92	GCGT	124	TCGT	156	ATGT	188	CTGT	220	GTGT	252	TTGT
29	ACTA	61	CCTA	93	GCTA	125	TCTA	157	ATTA	189	CTTA	221	GTTA	253	TTTA
30	ACTC	62	CCTC	94	GCTC	126	TCTC	158	ATTC	190	CTTC	222	GTTC	254	TTTC
31	ACTG	63	CCTG	95	GCTG	127	TCTG	159	ATTG	191	CTTG	223	GTTG	255	TTTG
32	ACTT	64	CCTT	96	GCTT	128	TCTT	160	ATTT	192	CTTT	224	GTTT	256	TTTT

4.3 Decryption

Decryption is a process of converting encrypted text back to original text. Only an authorized user can decrypt data because decryption requires a secret key or password.

The first block of decryption receives the encrypted message. The second block generate Pk value. The third block receives the encrypted DNA code and converts to its corresponding binary values. These pairs are substituted by 00 for A, 01 for T, 10 for G, and 11 for C. The fourth block the arrange binary values to block. The fifth block convert binary value to its Extended ASCII Value and the sixth block convert Extended ASCII value back to the original data or the decrypted message.

Step 1: The DNA sequences are separated from primers (start primer and end primer), since the message is in between them.

Step 2: The DNA nucleotides A, T, G, and C characters are substituted accordingly (00,01,11,10 respectively).

Step 3: Then they are converted into ASCII code and then the message is received at receiver end.

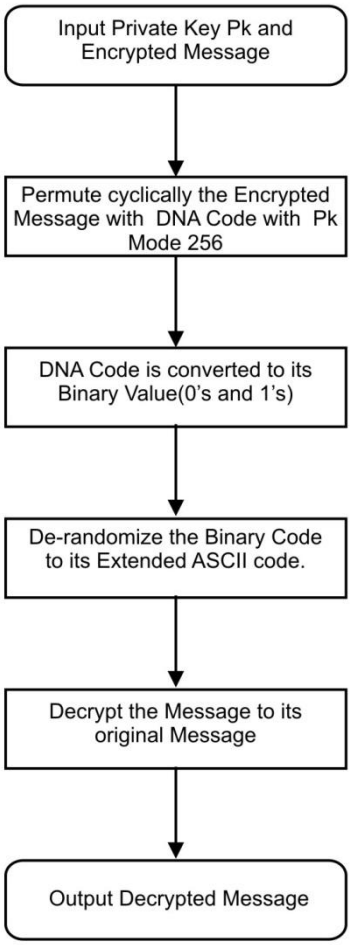

Fig 2: Decryption (Stage III of the Proposed system)

4.4 Example

The proposed system is explained with an example. Consider the source data be **BOY**

4.4.1 Encryption Process

Step 1: ASCII value of B= 66, O=79 , Y=89

B(66) → 01000010 → CAAG
O(79) → 01001111 → CATT
Y(89) → 01011001 → CCGC

Step 2: After DNA Combination substitution the message **BOY** will become **CAAG-CATT-CCGC.**

Step 3: From random key substitution table the DNA combination will become **35, 48, 58.**

Step 4: So the final encrypted message or cipher text is **35,48,58.**

4.4.2 3.4.2 Decryption Process

Step 1: Cipher text is 35,48,58.

Step 2: Substitute random generated key at an instance

35 → CAAG → 01000010

48 → CATT → 01001111

58 → CCGC → 01011001

Again

01000010 → 66 → B

01001111 → 79 → O

01011001 → 89 → Y

5. DISCUSSION

The proposed system is implemented using java as programming language and Window 7 as platform. Proposed system belongs to the modern symmetric key encryption system. Here each character is converted to an expanded set of characters. The specialty of the proposed system is that there is no need to use DNA chromosome or any other similar data during transmission, instead along with each character an inbuilt private key in the range of 1 to 256 will be transmitted.

The proposed system corresponds to the single character, two values are to be transmitted – one the encrypted integer value of Pk and the encrypted message. Based on the above discussion, one can conclude that security and the performance of the proposed algorithm are satisfactory for multi-level security applications of today's networks.

6. CONCLUSION

The proposed method of encoding is far better and faster than conventional cryptography like DES and other DNA based encryption algorithms. The proposal can be further enhanced to include in security mechanism of wireless networks and analyzing its performance to basic cryptanalytic attacks and comparing it with existing cryptosystems to know exactly how much improvement is achieved.

Security plays an important role in communication world. In order to protect the data from intruders, one need better security measures. The theoretical analysis shows that this method is more powerful against certain attacks. This method ensures confidentiality and data integrity over data transmission.

7. REFERENCES

[1] Adleman L M, Molecular Computation of Solutions to Combinatorial Problems, Science, Vol 266, pp 1021-1024, November 1994.

[2] K. Sireesha and V. Srujana, An overview and Analysis of Private & Public Key, *International Journal of Technological Exploration & Learning*, pp 281- 283, December 2013.

[3] Ayushi, A Symmetric Key Cryptographic Algorithm, *International Journal of Computer Applications* Vol 1, 2010.

[4] Bio-inspired computing: constituents and challenges, *International Journal of Bio-Inspired Computation*, Vol 1,pp 135-150, March 2009.

[5] Bonny B Raj and Panchami V, DNA Based Cryptography Using Permutation and Random Key Generation Method, *International Journal of Innovative Research in Science, Engineering and Technology*, Vol 3, pp 263-267, July 2014

[6] Gehani, T. LaBean, and J. Reif, DNA-Based Cryptography, *Lecture Notes in Computer Science, Springer*, Vol 2950, pp 167-188 2004.

[7] Rupali Soni and Gopal Prajapati, A Modern Review on Cryptographic Techniques, *International Journal of*

Advanced Research in Computer Science and Software Engineering Vol 3, Issue 7, pp162-167 July 2013.

[8] *http://www.blc.arizona.edu/molecular_graphics/dna_structure/dna_tutorial.html*

[9] Shipra Jain and Dr. Vishal Bhatnagar, Analogy of Various DNA Based Security Algorithms Using Cryptography and Steganography, International Conference on Issues and Challenges in Intelligent Computing Techniques pp 285-291, 2014.

[10] Anu priya Agarwal, Praveen Kanth, Secure Data Transmission using DNA Encryption, *Computer Engineering and Intelligent Systems* Vol 5, pp 51-59, 2014.

[11] "Introduction to Public-Key Cryptography", an article availabledeveloper.netscape.com/docs/manuals/security/pkin/contents.html.

[12] Olga Tornea, Monica Borda, Tatiana Hodorogea, and Mircea-Florin Vaida, Encryption System With Indexing Dna Chromosomes Cryptographic Algorithm, *Proceedings of the 7th IASTED International Conference*, pp12-1, February 2010.

[13] Sherif T. Amin , Magdy Saeb and Salah El-Gindi, A DNA-based Implementation of YAEA Encryption Algorithm.

[14] S. T. Amin, M. Saeb, S. El-Gindi, A DNA-base Implementation of YAEA Encryption Algorithm, *IASTEDInternational Conference on Computational Intelligence*,San Francisco, pp 120-125, 2006.

[15] Tatiana Hodorogea, Mircea-Florin Vaida, Alternate Cryptography Techniques, ICCC05, Miskolc-Lillafured, Hungary, Vol. 1, pp 513-518, 2005.

[16] R. K. Wilson, The sequence of Homo sapiens FOSMID clone ABC14-50190700J6, submitted to http://www.ncbi.nlm.nih.gov, 2009.