

(An ISO 3297: 2007 Certified Organization) Vol. 4, Issue 11, November 2016

# Cloud Data Security using DNA Cryptography Techniques

Vijay Prakash Tiwari, Akshay Bapusaheb Patil, Prof Avinash Murlidhar Ingole

B.E. Student, Dept. of Computer Engineering, Bharati Vidyapeeth's COE, Lavale, Pune, Maharashtra, India

B.E. Student, Dept. of Computer Engineering, Bharati Vidyapeeth's COE, Lavale, Pune, Maharashtra, India

Asst. Professor, Dept. of Computer Engineering, Bharati Vidyapeeth's COE, Lavale, Pune, Maharashtra, India

**ABSTRACT**: In this era, Data over the internet can found in various format like Text, Picture and Video, etc. Everyday data is generated with exponential rate, that leads to the situation where we need very precise, fast and accurate methods to keep them secure from possible threats. Various organizations as well as individuals are using Cloud Technology to share the information over the internet. This rises the security concerns when it comes to Cloud Systems. Today, there are many ciphering techniques available but theoretically most of them may fail once when there is requirement of processing huge information in real time. Using Molecular Science and Biotechnology, taking inspiration straight from the Mother Nature, DNA Computing can be used to solve this problem. DNA Cryptography Techniques, if successfully implemented in the nearby future, can result in the most secure environment to process information which even can be integrated with Cloud Technology and make data over cloud more Robust, Effective and Secure.

**KEYWORDS**: Cloud Computing, Cloud System, Cloud Storage, Cloud Data, Cloud Security, DNA Computing, DNA Cryptography Techniques, Encryption and Decryption, DNA, Cryptography

# I. INTRODUCTION

Cloud Computing is the field of computation where an organization or an individual stores data like text, picture or video over remotely hosted servers, rather than saving all of them on their own local storage machine or computer. Concept of Cloud Computing has been there since late 1970s in the form of Distributed Computing but popularized by Amazon.com on 2006. Currently, Cloud Technology is one of thewidely-used Computation Technology.

DNA Computing is the field of Molecular Computation which first was introduced by LeonardMax Adleman, an American Computer Scientist, in the year 2002[1]. DNA stands for Deoxyribonucleic Acid. DNA Stores the biological information of the living being about their various biological characteristics. DNA is made of several type of bases that is used for computational purpose. DNA Computation was introduced as a possible approach to solve Combinatorial Problems. The first problem that was solved using DNA Computing wasHamiltonian Graph[1]. The most important thing here to be noticed is that in DNA Computing, Computation is done using DNA not on DNA.

As we all are aware of that today, the data over the internet is increasing exponentially. There are various serious problems that is raised due to this that need to be addressed carefully. Among which, Security is one of the primary concern. Many Organization and Individuals are now using Cloud Based Storage Technology to save and share their data over the internet. Cloud System are secured using various Cryptographic Technologies like RSA, AES or HCC, etc. But these technologies may collapse in future when the data to be ciphered are huge in quantity.

DNA Computing can become handy when there is requirement to solve such gigantic problems related to ciphering the data because DNA has been proven as one of the robustciphered data storage device created by Mother Nature that effectively provides information to our Body and Brain about various activities, etc. Here we will discuss one of the possible way to implement the DNA based Cryptographic Technique to cipher the cloud based data and increase the Precision, Accuracy and reduce the processing time.



(An ISO 3297: 2007 Certified Organization)

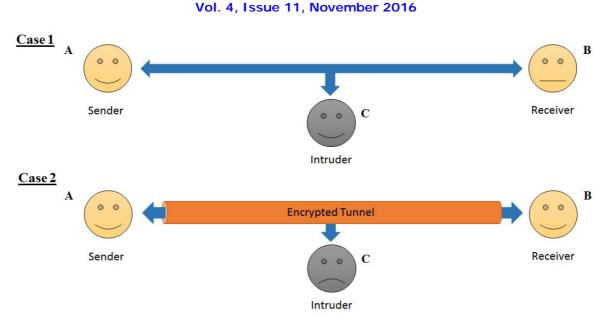


Fig 1: Non - Encrypted Channel vs Encrypted Channel

#### II. RELATED WORK

# A. CRYPTOGRAPHY

Cryptography is the study of techniques or methodology to encode the plain text into ciphered text and vice versa. Cryptography consist of basically two complementary sub techniques; i) Encryption ii) Decryption. Encryption is the technique to convert a plain text (understandable form) into ciphered/encrypted text (not understandable form) [5]. This process is known as Ciphering or Encrypting. Decryption is the technique to covert the ciphered/encrypted text (not understandable form) to plain text (understandable form). This is process is also known as Deciphering or Decrypting. Implementing Cryptography, makes the medium of communication secure and thus channel becomes a more reliable medium to send some secret data. This technique can briefly be understood by the Fig 1. In the Figure, Person 'A' (Sender) wants to send the message to Person 'B' (Receiver) but there is the Third Person 'C' (Intruder). In Case 1, there is no secure channel to transmit the data between Person 'A' and Person 'B' because the message that Person 'A' is sending, can also be understand by Person 'B' is a secure channel because the Person 'C' is getting Ciphered/Encrypted Data which is not understandable form of data because whatever the Person 'A' is sending, first it gets encrypted then sent on the channel than at the receiving end Person 'B' is getting that message by first decrypting it. Hence, the middle man or the intruder that is Person 'C' is getting an Encrypted Form or a Not Understandable Form of that message.

While Encryption - Decryption of a data, a secret key is required to perform the Encryption - Decryption. Encryption Key is required to encrypt the Plain Data and Decryption Key is required to decrypt the Encrypted Data. Cryptography can be broadly divided into two categories; i) Symmetric Cryptography ii) Asymmetric Cryptography [3]. Symmetric Cryptography consist of same secret key at both side, sender and receiver. This means that in Symmetric Cryptography the Encryption Key and Decryption Key are same. Asymmetric Cryptography uses two types of Secret keys; i) Private Key ii) Public Key. There are two types of Encryption Keys; i) Public Encryption key ii) Private Encryption Key. In Asymmetric Cryptography keeping the Private Key safe, we can assure that the data remain secured. Asymmetric Cryptography is time taking and require intensively complex processing but yet maintaining the security as maximum as possible. To make DNA based Cryptography, a more reliable but yet a fast medium to implement security we will use the Symmetric Cryptography Technique. To increase security further, we will use various existing Cloud Based System Security Techniques. There are various Cryptography implementing techniques but the core thing among all is that the degree of uncertainty and randomness in the process of generation of Secret Key. Higher the degree of



(An ISO 3297: 2007 Certified Organization)

## Vol. 4, Issue 11, November 2016

uncertainty and randomness, the more secure and stable medium we have but eventually increasing the processing requirement. Modern Cryptography uses various Mathematical Formulated Algorithm like Elliptical Curves, Matrix Manipulation, Prime Factorization, etc. while Quantum Cryptography utilizes the randomness of Quantum Particles or the state of electron of the atom. DNA based Cryptography utilizes the complexity and randomness of biological processes.

## B. DNA STRUCTURE AND BIO-LOGICAL THEOREM

DNA stands for Deoxyribonucleic Acid. DNA is molecule that is made up of 2 components. First, it has 4 types of Bases that are Thymine (T), Guanine (G), Cytosine (C) and Adenine (A). Second, DNA contains Sugar-Phosphate which makes its backbone as shown in Fig 2. DNA Bases are also known as Nucleotides. DNA consist of 2 biopolymerstrands and forms a double helix structure that is described in the Fig. 2. Between strands, lies Base Pairs that are bonded together with strong Hydrogen Bonding. This pair exist only in certain manner such that "Adenine (A) can only make Hydrogen Bond (double bond) with Thymine (T)" and "Cytosine (C) can only make Hydrogen Bond (triple bond) with Guanine (G)". The sequence of these base pair defines the rules for formation of Cell, eventually whole body of an organism. DNA is found in every cell of every living being. Basically, it is found in every nucleus of the cell and also in Mitochondria. It is a biological storage device that store all the genetic information and instructions which helps in formation of cells. DNA Molecule has two chemical polarity that is 5' and 3' at top and bottom as described in Fig 3. These two polarity enables DNA to bind together and transform itself into a double strand helix structure from single strand structure.

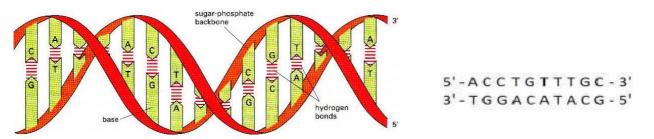


Fig 2: DNA Helix Structure

Fig 3: Example of DNA Molecular Orientation

DNA Molecule is responsible for transmitting massages among the cells. DNA uses proteins to interact with its environment. DNA uses mRNA (messenger Ribonucleic Acid) to send information. There are two processes that are involve in transmitting a message; i) Transcription ii) Translation. In Transcription, DNA passes the information to the mRNA. In Translation, mRNA uses the information to interact with proteins and passes on the desired message. DNA has the phenomenal biological property where it replicates without losing the original DNA Structural Information. Legitation Process is one of the process by which double strand structure of DNA form new double strand DNA.

#### III. DNA COMPUTING AND DNA CRYPTOGRAPHY

DNA Computing is the field of computing bringing together the Computer Science, Biological Science and Molecular Science to understand and solve some primary NP problem[1][2]. Earlier it was introduced by Leonard Max Adleman but now it has evolved as one of the most fascinating platform to develop something new by teachings of Mother Nature. DNA Computing is the best example of Biomolecular Computing. Biomolecular Computers are those computers where all the computing components are made up of Molecular Compounds i.e. all Input/Output and Software/Hardware are all in form of a Molecular Compound.

DNA Computing involves various steps such as Melting, Annealing, Merging, Amplification and Selection. DNA actually behave like a Turing Machine that is why it can be used as a Data Storage Device. Adleman has showed that DNA Computing can be used as an effective tool to solve the NP problems like Hamiltonian Graph Problem or Travelling Salesman Problem (TSP) [1]. He showed that DNA Computing can be used to solve complex Combinatorial Problems like TSP and Finite State Problem. Here, the basic idea is that all the operations are performed over DNA



(An ISO 3297: 2007 Certified Organization)

# Vol. 4, Issue 11, November 2016

(more precisely using DNA Bases or Nucleotide) not in DNA. DNA Computing can be classified as Intermolecular DNA Computing, Intramolecular DNA Computing and Supramolecular DNA Computing. DNA Computers form a Self-Replicating System.

TTTT	NUL	TTTC	Space	TTTG	@	TTTA	•	TTCT	Ç	TTCC	á	TTCG	L	TTCA	Ó
TTGT	SOH	TTGC	1	TTGG	А	TTGA	А	TTAT	ü	TTAC	í	TTAG	Т	TTAA	ß
TCTT	STX	TCTC		TCTG	В	TCTA	В	TCCT	é	TCCC	ó	TCCG	т	TCCA	Ô
TCGT	ETX	TCGC	#	TCGG	С	TCGA	с	TCAT	â	TCAC	ú	TCAG	ŀ	TCAA	Ò
TGTT	EOT	TGTC	\$	TGTG	D	TGTA	DEL	TGCT	ä	TGCC	ñ	TGCG	_	TGCA	õ
TGGT	ENQ	TGGC	%	TGGG	Е	TGGA	e	TGAT	à	TGAC	Ñ	TGAG	+	TGAA	Õ
TATT	ACK	TATC	&	TATG	F	TATA	f	TACT	å	TACC	a	TACG	ã	TACA	μ
TAGT	BEL	TAGC		TAGG	G	TAGA	g	TAAT	ç	TAAC	0	TAAG	Ã	TAAA	þ
CTTT	BS	CTTC	(	CTTG	Н	CTTA	h	CTCT	ê	CTCC	i	CTCG	L	CTCA	Þ
CTGT	TAB	CTGC	)	CTGG	Ι	CTGA	i	CTAT	ë	CTAC	R	CTAG	F	CTAA	Ú
CCTT	LF	CCTC	*	CCTG	J	CCTA	j	CCCT	è	CCCC	-	CCCG	<u>JL</u>	CCCA	Û
CCGT	VT	CCGC	+	CCGG	Κ	CCGA	k	CCAT	ï	CCAC	1⁄2	CCAG	ТГ	CCAA	Ù
CGTT	FF	CGTC	,	CGTG	L	CGTA	1	CGCT	î	CGCC	1⁄4	CGCG	ŀ	CGCA	ý
CGGT	CR	CGGC	-	CGGG	Μ	CGGA	m	CGAT	ì	CGAC	i	CGAG	=	CGAA	Ý
CATT	SO	CATC		CATG	Ν	CATA	n	CACT	Ä	CACC	<b>«</b>	CACG	÷	CACA	-
CAGT	SI	CAGC	/	CAGG	0	CAGA	0	CAAT	Å	CAAC	»	CAAG	¤	CAAA	1
GTTT	DLE	GTTC	0	GTTG	Р	GTTA	р	GTCT	É	GTCC		GTCG	ð	GTCA	
GTGT	DC1	GTGC	1	GTGG	Q	GTGA	q	GTAT	æ	GTAC		GTAG	Đ	GTAA	±
GCTT	DC2	GCTC	2	GCTG	R	GCTA	r	GCCT	Æ	GCCC		GCCG	Ê	GCCA	_
GCGT	DC3	GCGC	3	GCGG	S	GCGA	S	GCAT	ô	GCAC		GCAG	Ë	GCAA	3⁄4
GGTT	DC4	GGTC	4	GGTG	Т	GGTA	t	GGCT	ö	GGCC	+	GGCG	È	GGCA	ſ
GGGT	NAK	GGGC	5	GGGG	U	GGGA	u	GGAT	ò	GGAC	Á	GGAG	1	GGAA	§
GATT	SYN	GATC	6	GATG	V	GATA	v	GACT	û	GACC	Â	GACG	Í	GACA	÷
GAGT	ETB	GAGC	7	GAGG	W	GAGA	w	GAAT	ù	GAAC	À	GAAG	Î	GAAA	\$
ATTT	CAN	ATTC	8	ATTG	Х	ATTA	х	ATCT	ÿ	ATCC	©	ATCG	Ï	ATCA	0
ATGT	EM	ATGC	9	ATGG	Y	ATGA	у	ATAT	Ö	ATAC	╡	ATAG	L	ATAA	
ACTT	SUB	ACTC	:	ACTG	Ζ	ACTA	Z	ACCT	Ü	ACCC		ACCG	Г	ACCA	•
ACGT	ESC	ACGC	;	ACGG	[	ACGA	{	ACAT	ø	ACAC	٦	ACAG		ACAA	1
AGTT	FS	AGTC	<	AGTG	\	AGTA		AGCT	£	AGCC	Ц	AGCG		AGCA	3
AGGT	GS	AGGC	=	AGGG	]	AGGA	}	AGAT	Ø	AGAC	¢	AGAG	+	AGAA	2
AATT	RS	AATC	>	AATG	۸	AATA	~	AACT	×	AACC	¥	AACG	Ì	AACA	
AAGT	US	AAGC	?	AAGG	_	AAGA	DEL	AAAT	f	AAAC	٦	AAAG		AAAA	nbsp

#### Table 1: Genome-Encode Library Table

DNA Cryptography uses DNA Nucleotides only to generate a set of Symmetric Cryptographic Key. For, DNA Cryptography, many Techniques has already been established in many researches[4][6]but here, I aimed to develop a Technique to make the Existing Cloud-BasedData Storage Security Systems more accurate and giving the Encrypting and Decrypting capability directly to the AuthorizedClient on its own Machine. In this technique, first, we have to define three types of information. First of all, we need certain standard Library named as Genome-Encode Librarythat includes the 4-Bit Base Sequence uniquely defined for all 256 ASCII characters as shown in Table 1. This Genome-Encode Library is open for all to use and encode the given Plain Text message into Genome-Encoded Ciphered Text. Second we need Genome-Padding Library which will be use to replace the existing Genome Sequence to Padded Genome sequence. It is just to introduce an extra bit of randomnessto the data as described in Table 2. This Genome-



(An ISO 3297: 2007 Certified Organization)

# Vol. 4, Issue 11, November 2016

Padding Library is not standardized as it can be defined by the user itself. Third component we need is the Base-Binary Library that store the information about the equivalent conversion of Padded Genome Message to a long binary string i.e. Table 3.This Base-Binary Library is also not standardized as it can be defined by the user itself.

Α	ACGA
А	ACGA
С	TTGC
G	GCCT
Т	ATGC

Table 2: Genome-Padding Library Table Table 3: Base-Binary Library Table

#### IV. PROPOSED ALGORITHM

Aim of this proposed algorithm is to provided client end cryptography using DNA Computing which when integrated with existing Cloud System Storage Server Security Methods can increase reliability and secrecy of the data. The most important feature of this algorithm is that, it can provide not only textual data based Encryption-

Decryption Methodology but also provide the capability of Encrypting-Decrypting the Images, Audios and Videos.

Another important feature of this algorithm is that the data getting stored or data under transmission if even get hacked or intruded, that data will be of no use for the middle man even to that data administrator of the cloud storage facility.

# A. ENCRYPTION:

- Let Message be:M = "Hi!"
- Using Genome-Encode Library we can get that message M can be encode as follows (Using Table 1): M' = "CTTGCTGATTGC".

- Converting Binary String M''' into equivalent Decimal Value.
  M'''' = "164514448622248693223893078547".
- Send this data, M'''' to the Cloud Server to Store it.
- Generate the decryption key as: [Padded Value for A][Base-Binary for A][Padded Value for C][Base-Binary for C][Padded Value for G][Base-Binary for G][Padded Value for T][Base-Binary for T] Example: Decryption Key: - "ACGA11TTGC01GCCT10ATGC00".

# B. DECRYPTION:

- Download the file from the Cloud Server i.e. M''''' = "164514448622248693223893078547".
- Retrieve the decryption key i.e.
  Decryption Key: "ACGA11TTGC01GCCT10ATGC00".
- Separate the Padded Values and Base-Binary Values from it i.e.



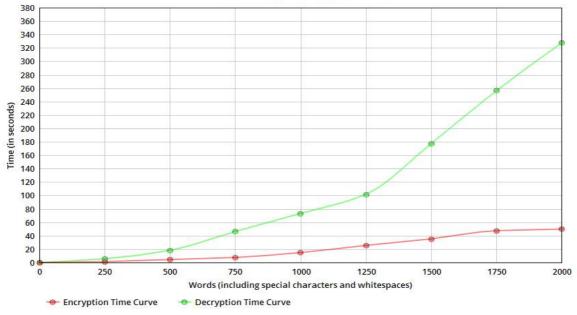
(An ISO 3297: 2007 Certified Organization)

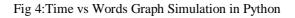
# Vol. 4, Issue 11, November 2016

- Using Padded Values, convert the Base String to corresponding Base String i.e. M' = "CTTGCTGATTGC".
- Using Genome-Encode Library, retrieve the original message i.e. M = "Hi!".

# V. SIMULATION AND RESULTS

## DNA Cryptographic Algorithm Execution





WORDS	ENCRYPTION TIME (RED CURVE)	<b>DECRYPTION TIME (GREEN CURVE)</b>
0	0	0
250	1.347783638	5.577723219
500	4.445763997	18.20145907
750	7.547622326	46.43985167
1000	14.88888552	73.19277255
1250	25.54612554	101.6897546
1500	35.23954941	177.4820321
1750	47.54051172	256.3900402
2000	49.89798045	327.56412

Table 4: Data Gathered after implementing the algorithm



(An ISO 3297: 2007 Certified Organization)

## Vol. 4, Issue 11, November 2016

Data in Table 4 is gathered using a Dell Inspiron 5545 Laptop that was running on Windows 10 with 8 GB RAM and 1TB Hard Disk. This Data is one of the worst-case analysis of the proposed algorithm. Since the Encryption and Decryption has to be done on the client's machine therefore this test was conducted on such machine.

Fig 4 describes the effectiveness and processing time taken by the proposed algorithm to encrypt and decrypt the same data. Here, the X-Axis represents the number of characters given as the input that also included the all the white spaces and other special characters. Y-Axis Represents the time taken by the algorithm proposed in seconds. Graphshown in Fig 4 represents that the algorithm is effective not only while extracting the data from cloud or uploading it on one of the cloud, it can be an effective when we need to perform a document based handshake or password based handshakes.

## VI. CONCLUSION AND FUTURE WORK

This research shows that it is just the beginning of the work. There are number of possibilities that the future holds when it comes to learning from Mother Nature and implementing those learnings to our existing technology to make it much precise and secure. DNA Cryptography when integrated with distributed computing technologies like Cloud Computing, Wearable Computing or Pervasive Computing, can result in a more reliable environment to save and share the data around. The proposed algorithm is one of the implementable aspect of this, that is also an underdevelopment technology. As many Cryptanalyst, has already said that the future of cryptography lies in the multidisciplinary studies of various aspects of science and Mathematics.

#### REFERENCES

- 1. Leonard Adleman, "Molecular computation of solutions of combinatorial problems", Science, Vol.266, 1994, pp. 1021-1024.
- Anup R. Nimje, "Cryptography in Cloud-Security Using DNA (Genetic) Techniques", International Journal of Engineering Research and Applications (IJERA) ISSN: 2248-9622 www.ijera.com Vol. 2, Issue5, September- October 2012, pp.1358-1359.
- 3. Neha Pallavi\*, Archana Singh and Surya Prakash Dwivedi"A DNA Based Secure Data Hiding Technique for Cloud Computing", International Journal of Current Engineering and Technology 11 July 2016, Vol.6, No.4.
- 4. Anchal Jain and Navin Rajpal, "Adaptive Key Length Based Encryption Algorithm using DNA Approach", International Conference on Machine Intelligence Research and Advancement.
- Snehal Javheri and Rahul Kulkarni, "Secure Data communication and Cryptography based on DNA based Message Encoding", International Journal of Computer Applications (0975 – 8887) Volume 98– No.16, July 2014.
- 6. Noorul Hussain UbaidurRahmana, Chithralekha Balamuruganb, Rajapandian Mariappanc, "A Novel DNA Computing based Encryption and Decryption Algorithm", International Conference on Information and Communication Technologies.
- Li Xin she, Zhang Lei, Hu Yu pu. A Novel Generation Key Scheme Based on DNA. In: Proceedings of the International Conferenceon Computational Intelligence and Security; 2008.p. 264-266.
- Mona Sabry, Mohamed Hashem, Taymoor Nazmy. Three Reversible Data Encoding Algorithms based on DNA and Amino AcidsStructure. International Journal of Computer Applications 2012; 54: 0975 – 8887.
- 9. NRDC, Govt. of India, http://www.nrdcindia.com/Patent%20Asistance%20 (in%20India) %20Form%202011.pdf.
- 10. O Tornea, ME Borda. DNA Cryptographic Algorithms. In: IFMBE Proceedings of the International Conference on Advancements of
- 11. Medicine and Health Care through Technology: 2009 Sep 23-26; Cluj-Napoca, Romania. Springer; 2009. p 223-226.
- 12. Padma Bt. DNA computing theory with ECC' http://www.scribd.com/doc/55154238/Report, 2010.
- 13. Qiang Zhang, Ling Guo, Xianglian Xue, Xiaopeng Wei. An image encryption algorithm based on DNA sequence addition operation.
- 14. Vijay Prakash Tiwari, Vikas Tiwari and U. C. Patkar, "DNA Computing and Its Implementations", International Journal on Recent and Innovation in Trends in Computing and Communication, ISSN: 2321-8169, Volume 4 issue 4.

# BIOGRAPHY

Vijay Prakash Tiwari is a Computer Engineer Student of Bharati Vidyapeeth's College of Engineering, Lavale, Pune, Maharashtra, India. He has published 6 review papers including "DNA Computing and Its Implementations" in IJRITCC Volume 4 Issue 4, "Li-Fi Technology, Implementation and Applications" in IRJETVolume 3 Issue 4, "Blue Brain Technology to Preserve Intelligence" in IRJET Volume 3 Issue 4, "Face Analysis for Improvised System Security using LASER Focus" in IRJET Volume 3 Issue 2, "Wearable Technologies and Extended Applications" in IJRITCC Volume 4 Issue 4 and "Augmented Reality and Its Technology" in IRJET Volume 3 Issue 4.

Akshay Bapusaheb Patil is a Computer Engineer Student of Bharati Vidyapeeth's College of Engineering, Lavale, Pune, Maharashtra, India. He has published 2 review papers including "Li-Fi Technology, Implementation and Applications" in IRJET Volume 3 Issue 4 and "Blue Brain Technology to Preserve Intelligence" in IRJET Volume 3 Issue 4.

Prof Avinash Murlidhar Ingole is Assistant Professor in Computer Engineer Department of Bharati Vidyapeeth's College of Engineering, Lavale, Pune, Maharashtra, India.