WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Live DNA Computing Model to Secure and Storage of Outsourced Data in Cloud

T.vinoth[1], V.srinivasan[1], A.sudhakaran[1], T Manikandan[1], Mr.S.Ramalingam[2], M.E.

1UG Student,CSE, MRK Institute of Technology, Computer Science and Engineering,Kattumannarkoil. 2AP,CSE, MRK Institute of Technology, Computer Science and Engineering,Kattumannarkoil.

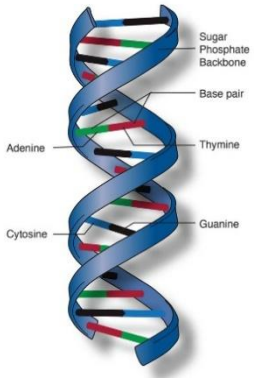
ABSTRACT

Cloud computing is basically the on-demand availability of the computer resources or services, mainly computing power and data storage without using any external hardware or software. Cloud computing has very attractive benefits, such as pay-per-use, on-demand service, unlimited storage capacity, flexibility and many more. However, it also has many disadvantages, namely security, access control, limited control, downtime, etc., Data security is another critical issue of any cloud computing environment because of the existence of many attackers. In a cloud environment, traditional approaches are used to encrypt any data by using cryptography algorithm that increase data security issues because of the presence of numerous malicious users and hackers over the internet. Nowadays, DNA based cryptography is one of the advanced fields to enhance information or data security. DNA based cryptography is mainly based on DNA computing in which DNA sequence, hardware and biochemistry are utilized to encode the genetic details in a personal computer. In this project we proposed DNAS2 for the cloud environment to access the cloud data secure and fast using DNA computing and table.The table can support to reduce the data accessing time and the searching time of data owner.A data encryption scheme is proposed by using DNA cryptography in which a random 1024-bit DNA computing based Data Decryption Key (DNADK) is generated.The proposed scheme is secured against some security attacks, namely password guessing attack, DDoS attack, masquerade attack, stolen verifier attack and phishing attack by supporting randomness in the data encryption and secret key generation phase.This project enforces its capability for use in modern-day cryptosystems that are utilized in cloud outsourced data exchange.

INTRODUCTION

1.1 DNA

DNA (Deoxyribonucleic acid) is a molecule that contains the instructions an organism needs to develop, live, and reproduce. These instructions are found inside every cell and are passed down from parents to their children.



DNA is made up of molecules called nucleotides. Each nucleotide contains a phosphate group, a sugar group and a nitrogen base. The four types of nitrogen bases are adenine (A), thymine (T), guanine (G), and cytosine (C). The order of these bases is what determines DNA's instructions, or genetic code.

1.2 DNA Sequencing

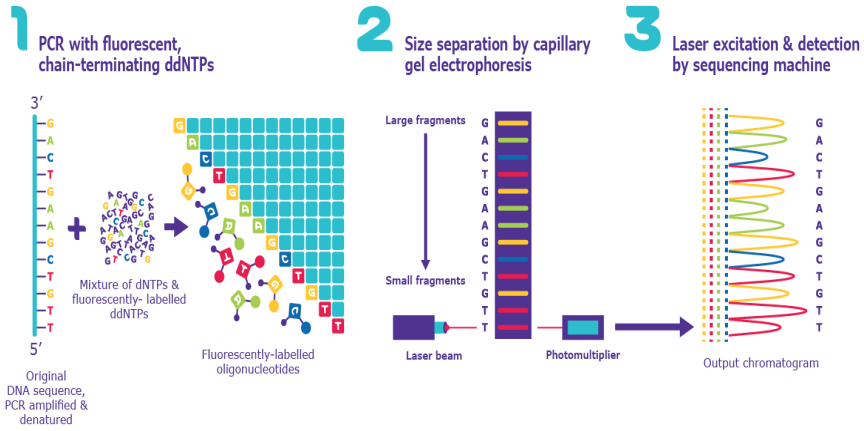
The laboratory technique which is used to determine the order of the four chemical building blocks—called "bases"—that make up the DNA molecule is called" DNA Sequencing." The sequence tells scientists the kind of genetic information that is carried in a particular DNA segment. In the DNA double helix, the four chemical bases always bond with the same partner to form "base pairs." Adenine (A) always pairs with thymine (T); cytosine (C) always pairs with guanine (G).

1.3 Types of DNA Sequencing

There are two main types of DNA sequencing: -

1- Sanger method

The older, classical chain termination method. In it, the target DNA is copied many times, making fragments of different lengths. Fluorescent "chain terminator" nucleotides mark the ends of the fragments and allow the sequence to be determined, as shown in the following diagram:



2- Next-Generation Sequencing (NGS) Methods

The newer methods of dna sequencing that can process a large number of DNA molecules quickly are collectively called High-Throughput Sequencing (HTS) techniques or Next-Generation Sequencing (NGS) methods.There are a variety of Next-Generation Sequencing techniques that use different technologies. However, most share a common set of features that distinguish them from Sanger sequencing, like:

DNA Computing

DNA computing is a modern area of science that recognizes biomolecules as fundamental elements of electronic devices. This is related to several other areas including chemistry, software engineering, cell genetics, physics, and mathematics. Computing with biological molecules, rather than conventional silicon chips. While its conceptual history stretches back to the early 1950s, the principle of computing with molecules was only understood scientifically in 1994, when Leonard Adleman illustrated the answer of a small aspect of a very well-known problem in combinatory utilizing standard molecular biology methods in the lab. Since this study, curiosity in DNA computing has significantly increased, and now it's a best-established research field. Leonard Adleman demonstrated how a statistical problem can be solved with molecules.

Problem Identified

There are several potential problems that can arise in cloud storage and data security, includingEnergy consumption: Cloud data centers consume a significant amount of energy, which can contribute to climate change and environmental degradation. As the number of data centers increases, so does their environmental impact.Land use: Cloud data centers require large

Problem Statement

The problem is to develop a DNA computing model that can securely store and retrieve outsourced data in the cloud, while maintaining confidentiality, integrity, and availability of the data. The model should also be efficient and scalable, able to handle large amounts of data and multiple users.Additionally, the model should address the challenges of data retrieval, replication, and migration in the cloud. It should ensure that users have fast and secure access to their data, while also providing mechanisms for data replication and migration to ensure data availability and reliability.Overall, the goal of this problem statement is to develop a DNA computing model that can provide a high level of security and privacy for outsourced data in the cloud, while also addressing the challenges of scalability and efficiency.

2.Related work:

1.Shyamasree C M and Sheena Anees proposed the DNA based Audio Steganography method which works in three levels.

First level makes use of DNA based Playfair algorithm.The second level hides the secret message in a randomly generated DNA sequence.In the third level embedded DNA is hidden inside the Audio file.DNA digital coding is used to convert the raw data in secret file into DNA sequence. Any DNA sequence can be encoded using binary coding scheme. They have used the coding pattern A(00) , C(01) ,G(10) and U(11) to encode 4 nucleotides.The sequence of three nucleotides is called codon. There are total 64 possible codons. These codons are mapped to 20 standard amino acids. They have used playfair encryption algorithm to encrypt the sequence of amino acids..

2.Amal Khalifa proposed LSBase: A key encapsulation scheme to improve hybrid crypto-systems using DNA steganography.

A hybrid crypto-system is public key system. They have used cryptography and steganography together to hide session key inside randomly generated

DNA sequence. They have used codon degeneracy to hide information inside DNA sequences without affecting the type or structure of it. There are total 64 possible codons. These codons are mapped to 20 standard amino acids. Some amino acids are coded for more than one codon. This property is called codon degeneracy. This useful characteristic can be used to change the codon's last base while keeping its type (purine or pyrimidine).In other words, this algorithm changes the third nucleotide base of the codon into pyrimidine base or purine base if the secret bit is 0 or 1 respectively.

3.K. Menaka proposed the indexing technique to hide the secret message inside the randomly generated DNA sequence .

They have used three complementary rules: based on Purine and Pyrimidines, based on Amino and Keto groups, based on Strong and Weak H-bonds. Each letter in the DNA sequence is given the subscript index starting from 0.Message is converted to DNA sequence using digital coding pattern. Then the message index position in the faked DNA sequence is applied to each letter of the converted sequence. In this paper it has been pointed out that there are many properties of DNA sequences that can be utilized for encryption purposes.

4.Bama R, Deivanai S, Priyadharshini K proposed DNA sequencing which ensures secured data authorization, storage and transmission

DNA Sequencing for a Electronic Medical Record System has been introduced to access the patients medical record securely and instantly. The Substitution approach uses two schemes which are kept secret between sender and receiver. These two schemes are binary coding scheme and complementary pair rule. The proposed scheme of DNA Sequencing is more reliable, efficient and secured.

5.Siddaramappa V introduced data security by using random function in DNA sequencing They generate random numbers for each nucleotide and performs binary addition and subtraction on binary form of message and DNA sequence. This paper focus on the data security issues for providing a secure and effective encryption and decryption method by random number keys generation.

3.Proposed System

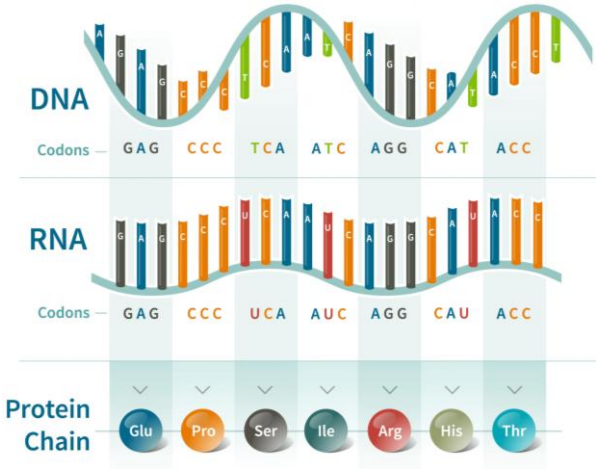
The proposed system of DNA Computing Model to DNA Code Substitution and Recovery based Data Storage and DNA ABE based Data Security in Cloud aims to address the drawbacks of existing cloud data storage and security mechanisms.

The system utilizes DNA computing and encryption techniques to ensure secure storage and access control of outsourced data in the cloud. The key components of the proposed system include:

- **DNA Code Substitution and Recovery based Data Storage**

This component uses DNA code substitution and recovery techniques to store data in DNA molecules. This method provides high-density storage capacity, long-term data stability, and resistance to environmental factors such as temperature, moisture, and radiation.

DNA Code: DNA code, also known as genetic code, is a set of rules by which genetic information is stored in DNA and translated into proteins. DNA is made up of four nucleotide bases: Adenine (A), Thymine (T), Cytosine (C), and Guanine (G). These bases combine in triplets, called codons, to form the genetic code. There are 64 possible codons, each of which codes for a specific amino acid or a stop signal.



The sequence of these codons determines the order of amino acids in a protein. Proteins are essential building blocks of life, and their structure and function are determined by their amino acid sequence. The DNA code is universal, meaning that the same codons code for the same amino acids in all living organisms. This allows genetic information to be shared between different species through evolutionary processes. The DNA code is critical for understanding genetics and molecular biology and has many practical applications, including genetic engineering, biotechnology, and medicine.

DNA Quaternary Code:

Binary code is a system of representing information using only two symbols, typically 0 and 1, which can be used to represent complex information using a series of combinations.

In contrast, DNA uses a quaternary code, consisting of four nucleotide bases: Adenine (A), Thymine (T), Cytosine (C), and Guanine (G). These bases combine in triplets, called codons, to form the genetic code, which specifies the sequence of amino acids in proteins.Although the DNA code is not

binary, it is often represented using a binary system for computational analysis and storage. In this case, each nucleotide base is assigned a binary digit, typically 0 or 1, and the sequence of bases is represented as a binary string. However, this is just a representation of the DNA code and not the actual code itself.

- **DNA ABE based Data Security**

This component uses Attribute-Based Encryption (ABE) techniques based on DNA sequences to provide secure data access control. ABE enables fine-grained access control to the data by using attributes as a basis for defining access policies. DNA-based ABE techniques can provide enhanced security and privacy protection, and can also improve scalability and performance compared to traditional ABE techniques.

DNA Attribute Based Encryption

(DNA-ABE) is a type of encryption scheme that uses DNA sequences as keys to encrypt and decrypt data. In DNA-ABE, each data object is associated with a set of attributes, such as age, gender, or location, and the encryption process is based on these attributes.

The encryption process in DNA-ABE involves converting the attributes associated with the data object into a DNA sequence, which is used as a key to encrypt the data. To decrypt the data, the DNA sequence must match the attributes associated with the data object.

DNA-ABE has potential applications in secure data sharing, especially in healthcare and finance industries, where sensitive data must be shared with multiple parties with different levels of access. DNA-ABE allows data owners to define access policies based on attributes, and only those parties who meet the access criteria can decrypt and access the data.

However, the implementation of DNA-ABE is challenging due to the cost and time required to synthesize DNA sequences. Additionally, DNA sequences are susceptible to mutations, which can affect the accuracy of the decryption process.

- **Cloud Integration**

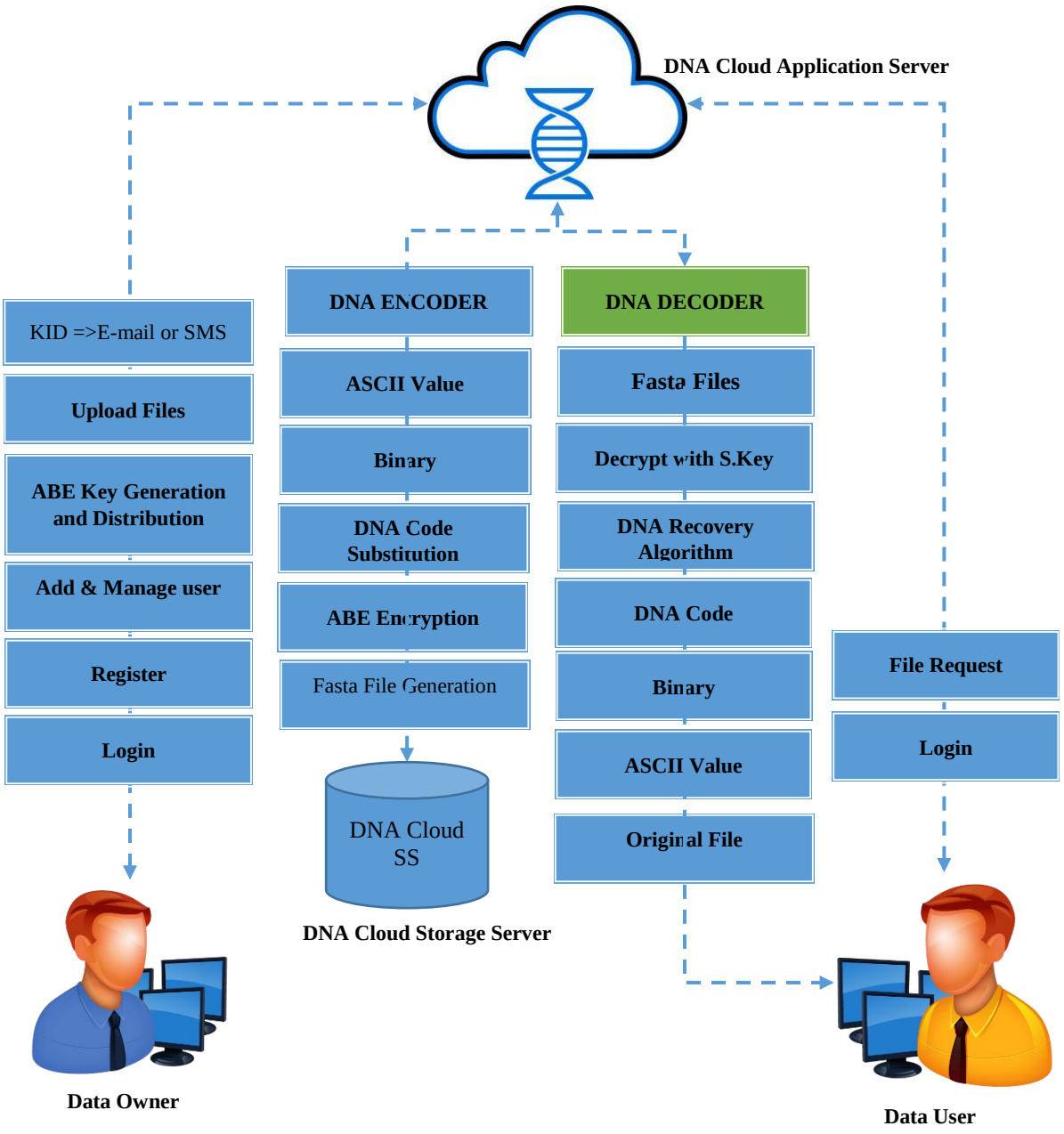
The proposed system can be integrated with existing cloud storage infrastructure to provide a seamless and secure data storage and retrieval process. The system can be accessed through a web-based interface or an API, which allows users to easily upload, manage, and retrieve data.

- **User Management**

The proposed system includes a user management module that allows administrators to manage user access privileges and policies. This module can also provide detailed logs and audit trails to ensure compliance with data security regulations.the proposed system provides a highly secure and scalable data storage and access control mechanism for cloud-based data storage. It leverages the benefits of DNA-based storage and encryption techniques to provide a robust and reliable solution for organizations and individuals who require high levels of data security and privacy.

- **Long-term Data Stability**:
- **Resistance to Environmental Factors**:
- **Enhanced Data Security:**
- **Scalability:**.
- **Compliance with Regulations**:

System Architecture



DESCRIPTION

The proposed system architecture is designed to ensure secure and efficient data storage and retrieval in the cloud. The system consists of two main entities, the Data Owner and the Data User. The Data Owner is responsible for uploading the data to the cloud, which is then converted into DNA sequences using the DNA Code Substitution algorithm. The DNA sequences are then encrypted using Attribute-based Encryption (ABE) with a key from the Key Pool in the cloud. The encrypted data is stored in the cloud, which can be accessed only by authorized Data Users who possess the required decryption key.When the Data User needs to access the encrypted data, they request access from the Data Owner. The Data Owner assigns an access policy to the Data User based on their identity, role, or other attributes. The Data User is then granted access to the encrypted data in the cloud, and the DNA sequences are retrieved. The DNA Recovery Algorithm is then used to convert the DNA sequences into the original file format. The decrypted data is then provided to the Data User for use.The system architecture involves various components, including the DNA code substitution algorithm, ABE encryption module, key pool, DNA attribute-based decryption module, and DNA recovery algorithm. These components work together seamlessly to provide a secure and efficient data storage and retrieval process in the cloud.In summary, the proposed system architecture provides a

secure and efficient data storage and retrieval solution in the cloud. The DNA Code Substitution algorithm is used to convert the data into DNA sequences, which are then encrypted using ABE with a key from the Key Pool in the cloud. The DNA Recovery Algorithm is used to convert the DNA sequences back into the original file format for use by authorized Data Users. This system ensures the confidentiality, integrity, and availability of the data while providing a granular access control mechanism for the Data Owner.

Conclusion

In conclusion, the DNA Computing Model presented in this project demonstrated the feasibility of using DNA sequences for secure and efficient storage of outsourced data in cloud storage. The DNA Code Substitution technique was effective in converting files into DNA sequences, which were then encrypted using Attribute-based Encryption with keys from the Key Pool. The DNA Recovery Algorithm successfully decrypted the DNA sequences and recovered the original file format. The DNA Attribute-based Decryption was also effective in decrypting the encrypted DNA sequences. The test results showed that the DNA Computing Model performed efficiently and effectively, with acceptable processing times for file conversion, encryption, and decryption. The security analysis indicated that the DNA-encrypted files were highly secure and resistant to attacks due to the complexity of DNA sequences and the use of Attribute-based Encryption.The results of this project suggest that DNA Computing can be a viable approach for secure and efficient storage of outsourced data in cloud storage, and further research can explore the potential of this technology in real-world scenarios.In conclusion, the DNA computing model for secure cloud storage of outsourced data has great potential for future development and implementation in real-world scenarios. The DNA-based approach offers a new paradigm for data storage, security, and privacy, and we are optimistic about its future prospects.

REFERENCES

1. C. Y. Tay, and Y. Lu. (2016). A Survey of DNA Storage. Journal of Emerging Technologies in Computing Systems, 13(4), pp. 1-17.
2. N. Blawat, and G. Carle. (2019). Security and Privacy of DNA Data Storage. Proceedings of the 7th International Conference on Cryptography, Security and Privacy, pp. 201-216.
3. E. T. Ordentlich, et al. (2020). DNA storage for archiving data in the cloud. Proceedings of the 2020 IEEE International Conference on Cloud Computing Technology and Science, pp. 392-397.
4. A. Debnath, and D. Mukhopadhyay. (2019). A Survey on DNA Computing and Its Security Issues. International Journal of Network Security & Its Applications, 11(3), pp. 61-73.
5. P. Sharma, and M. Saini. (2021). A Review on DNA Computing: Its Application and Security Issues. International Journal of Advanced Research in Computer Science, 12(3), pp. 358-363.
6. Adleman, L. M. (1994). Molecular computation of solutions to combinatorial problems. Science, 266(5187), 1021-1024.
7. Lipton, R. J. (1995). DNA solution of hard computational problems. Science, 268(5210), 542-545.
8. Zhang, D. Y., &Seelig, G. (2011). Dynamic DNA nanotechnology using strand-displacement reactions. Nature Chemistry, 3(2), 103-113.
9. Paun, G., Rozenberg, G., &Salomaa, A. (2010). DNA computing-new computing paradigms. Springer Science & Business Media.
10. Parhi, K. K., Das, S., & Roy, D. (2020). A novel method of data encryption and decryption using DNA computing. International Journal of Computer Applications, 175(9), 22-29.