Manipal Journal of Science and Technology

Volume 8 | Issue 2 Article 5

12-31-2023

Human Genome Analysis and Encryption: A Cloud Prototype Model Using AWS and AES

Priyanka Kunte
priyanka.mitmpl2023@learner.manipal.edu

Rachana Shanbhogue
Sahyadri College of Engineering and Management, rachana20015@gmail.com

Sanjana S. Nayak
Sahyadri College of Engineering and Management, nayaksanjana14@gmail.com

Neelima Bayyapu
Manipal Institute of Technology, neelima.bayyapu@manipal.edu

Manjunath K N
Manipal Institute of Technology, manjunath.kn@manipal.edu

See next page for additional authors

Human Genome Analysis and Encryption: A Cloud Prototype Model Using AWS and AES

Authors

Priyanka Kunte, Rachana Shanbhogue, Sanjana S. Nayak, Neelima Bayyapu, Manjunath K N, and Neha S. Gandhi

Research Article

Human Genome Analysis and Encryption: A Cloud Prototype Model Using AWS and AES

Priyanka Kunte, Rachana Shanbhague, Sanjana S Nayak, Neelima Bayyapu, Manjunath K N, Neha S Gandhi*

Email: priyanka.mitmpl2023@learner.manipal.edu, nayaksanjana14@gmail.com, rachana20015@gmail.com, neelima.bayyapu@manipal.edu, k.moorthi@manipal.edu, manjunath.kn@manipal.edu, neha.gandhi@manipal.edu

Abstract

SecureDNA presents a robust solution for safeguarding genetic data in cloud storage through the implementation of AES-256-CBC encryption. This innovative system, designed with a user-friendly interface, guides users seamlessly through DNA data upload processes, ensuring confidentiality and integrity. Utilizing AWS services such as RDS, VPC, and EC2, SecureDNA establishes a secure infrastructure for data storage and user interaction. By encrypting DNA records before storage and leveraging cloud computing benefits, this system not only protects sensitive genetic information but also offers scalability and accessibility. SecureDNA stands as a pivotal advancement in ensuring the privacy and reliability of genetic data in the digital age, with implications for healthcare, genetic research, and forensic sciences. This comprehensive approach addresses the pressing need for secure handling of genetic information, providing a crucial framework for future data protection standards. Through SecureDNA, users can confidently engage with cloud-based genetic data, knowing that their information is shielded from unauthorized access, contributing to advancements in personalized medicine and genetic research worldwide.

Keywords: AES, Cloud Storage, Cryptography, Data Security

1. Introduction

Cloud storage is a revolutionary technology in cloud computing where data is stored on the internet and managed by a cloud computing provider. This approach enables seamless transmission, storage, files and data retrieval of files and data remotely. However, ensuring the security of this data poses a significant challenge in cloud computing. Storing or transferring files and data to cloud systems increases the risk of physical access threats. Unfortunately, some cloud vendors neglect proper security regulations when storing client data. Data without the knowledge of those involved in communication. On the other hand, active attacks give attackers the ability to manipulate or delete data and resources within our system.

The vulnerabilities in a cloud network can have serious consequences, including financial losses and damage to the platform's reputation, especially if it serves a large public audience. Hence, there is a strong push for adopting new solutions that address these security concerns. To tackle these issues, cryptography algorithms like the Advanced Encryption Standard (AES) play a vital role. Cryptography works by transforming information into an unreadable form, making it inaccessible to unauthorized users who may try to understand or access the original content of files or information.

Priyanka Kunte, Sanjana S Nayak, Rachana Shanbhague, Neelima Bayyapu, Manjunath K N, Neha S Gandhi*

Computer Science and Engineering, Manipal Institute of Technology

Manuscript received: 18-10-2023

Revision accepted: 29-11-2023

* Corresponding Author

How to cite this article: Priyanka Kunte, Rachana Shanbhague, Sanjana S Nayak, Neelima Bayyapu, Manjunath K N and Neha S Gandhi, "Human Genome Analysis and Encryption", *Manipal J. Sci. Tech*., vol.8(2), 30-35, 2023.

The AES is a cryptographic algorithm endorsed by the Federal Information Processing Standards (FIPS) for securing electronic data. AES functions as both an encryptor and decryptor, using a symmetric block cipher. The US government has adopted AES, and it is now used worldwide. This algorithm operates as a symmetric key algorithm. Deoxyribonucleic acid (DNA), which carries essential genetic information, relies on bases such as adenine, guanine, cytosine, and thymine. Encryption acts as a guardian, preventing any unauthorized tampering. Applying encryption to DNA augments genetic privacy, with implications for various fields such as forensics, offering protection against the misuse of sensitive information. DNA encryption occurs on a website; subsequently, the entire website is stored on the cloud, enabling efficient and flexible DNA encryption and decryption processes. Storing the website on the cloud offers advantages such as scalability, accessibility, cost-effectiveness, data redundancy, and simplified maintenance.

Objectives:

- Create a secure web application for DNA sequence input.
- Use AWS cloud services, Relational Database Service (RDS) for managing the database, Elastic Compute Cloud (EC2) for the web application hosting, and Virtual Private Cloud (VPC) for network isolation and security.
- Establish strong authentication (registration and login) for users.
- Employ AES-256-CBC encryption to secure user information and DNA details.
- Collect user data and DNA sequences after administrator authentication.
- Safely store data in the cloud with encryption for increased security.
- Allow analysis of stored data for different genetic disorders.

2. Literature Review:

The evolution of secure file storage in cloud environments is marked by a constant quest for robust encryption methods and pioneering approaches to data security. This literature review delves into a collection of studies and research works that play a pivotal role in the ongoing discussion on secure data storage. Our exploration includes the use of hybrid cryptography, DNA-based encryption techniques, and biometric authentication. These emerging technologies offer promising ways to improve the confidentiality, integrity, and accessibility of sensitive data, particularly in the field of genetic analysis.

Title	Year	Author	Conclusions
Secure File Storage on the Cloud Using Hybrid Cryptography	2023	Smith, B Johnson, C Lee	Utilizes a hybrid approach with symmetric and asymmetric encryption for enhanced cloud security in file storage.
DNA Cryptography for Secure Data Storage in the Cloud	2018	Sukumaran S C, Mohammed M	Uses DNA sequences as cryptographic keys for secure data storage in cloud environments, particularly useful in fields like forensics.
Multi-Level DNA Encryption Technique Based on DNA Arithmetic and Biological Operations	2018	Zebari D A, Haron H	Introduces a multi-level DNA encryption technique combining DNA arithmetic and biological operations for enhanced security in DNA-based encryption for cloud storage.
Analysis of DNA Sequence Classification Using CNN and Hybrid Models	2021	Gunasekaran H, Ramalakshmi K, Arokiaraj A R M, Kanmani S D, Venkatesan C, Gnana Dhas C S	Discusses using CNN and hybrid models to analyze DNA sequences, potentially useful for DNA encryption and storage.
Enhanced DNA Cryptosystem for Secure Cloud Data Storage	2021	Sudarshan S, Thangavel M, Sowmiya B, Abhijith V S, Varalakshmi P	Proposes an enhanced DNA cryptosystem designed for secure cloud data storage, aiming to improve the efficiency and reliability of DNA-based

Title	Year	Author	Conclusions
			encryption methods in cloud environments.
Secure File Storage in Cloud Computing Using Hybrid Cryptography Algorithm	2016	Maitri P V, Verma A	Discusses a hybrid cryptography algorithm for secure file storage in cloud computing, enhancing security with symmetric and asymmetric encryption methods.
A Survey on Biometric Authentication: Toward Secure and Privacy-Preserving Identification	2018	Rui Z, Yan Z	Provides an overview of biometric authentication methods focusing on secure and privacy-preserving identification, relevant for access control in DNA encryption/storage systems.
Design of DNA-Based AES	2015	Sabry M, Hashem M, Nazmy T, Khalifa M E	Focuses on the design of DNA-based AES, exploring DNA sequences as cryptographic keys for AES encryption, potentially useful for secure data storage.
Cloud-Oriented Distributed and Encrypted File Storage (CODE-FS)	2018	Manek M, Chhadwa A, Shah K, Potey M, Khan M	Presents a cloud-oriented distributed and encrypted file storage system, discussing strategies for securely storing and accessing files in cloud environments, relevant for DNA data storage.

3. Proposed Methodology:

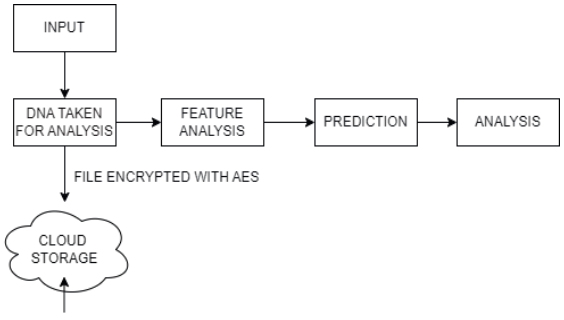

Figure 1: Secure DNA Web Application Architecture on Cloud

This work proposes securing data for cloud storage using an AES-based encryption technique (Figure 1).

The design of the architecture focuses on handling a variety of tasks, including user authentication, data input, and storage. Users are led through a process that commences with the application server. The administrator holds a crucial role in the system's operation and needs authorized access to kickstart the procedure by logging into the website. In cases where a user account is absent, the system guides them to register, thereby forming a secure and verified user community. Upon successful registration, users can sign in using their unique credentials and avail themselves of all features provided by the system.

Once users successfully log in, they are greeted with a detailed dashboard on the website. This dashboard allows them to easily upload their DNA data, a critical step in unlocking genetic insights. The user-friendly interface is crafted to guide users seamlessly through this process. Each DNA input generates a new database entry, contributing to the development of a vast repository of user data.

The thoroughness of this web application also encompasses data security and availability. Upon successful entry of DNA data, encryption plays a vital role in securing the information before storage. This measure guarantees that confidential genetic data remains safeguarded against unauthorized access. The encrypted data is then stored securely in cloud storage using cutting-edge cloud computing technology for improved accessibility.

In the web application architecture, there is a heavy reliance on the services offered by AWS. Specifically, the use of Relational Database Service

(RDS), Virtual Private Cloud (VPC), and Elastic Compute Cloud (EC2) is made to host the website and securely store data in the cloud.

RDS - a managed database service that helps us set up and handle our relational database. This database holds crucial information like user details and DNA sequences. RDS provides features such as automated backups, maintenance support, and compatibility with various database engines such as MySQL and PostgreSQL. This ensures that our data always stays safe and accessible.

Moving on to VPC, it plays a vital role in our architecture by creating an isolated virtual network environment within the AWS cloud. With VPC, we have control over network settings like IP address ranges and access controls. This in turn empowers us to establish a secure private network for our website and database. We enhance security against unauthorized access and potential threats by isolating our resources.

Now onto hosting the web application itself - here is where EC2 instances come into play. These instances function as virtual servers in the cloud handling the backend operations of our application. Users engage with our website through these instances, inputting their DNA data and exploring different features. EC2 offers flexibility in terms of instance types so we can adjust resources based on demand ensuring optimal performance even during peak user traffic.

By harnessing the capabilities of RDS, VPC, and EC2 together, we craft a robust infrastructure for our web application. This setup not only guarantees data integrity and accessibility but also provides scalability and flexibility essential for meeting the evolving needs of our users.

Regarding safeguarding data in cloud storage, our approach involves using the AES-256-CBC encryption method to uphold the privacy and reliability of delicate DNA details. The first step initiates with retrieving the encrypted DNA records from the database, which are housed within the `$row ['DNA']` variable.

Following this retrieval, we define the encryption cipher as `"AES-256-CBC"` and allocate a private encryption key of `"12345678901234567890123456789012"` to the `$secret` variable. This 32-character secret key plays a vital role in both encrypting and decrypting procedures and must be handled with utmost care.

To pave the way for decryption, we configure the encryption settings to their default state of `0`. Subsequently, an initialization vector (IV) is generated for AES-256-CBC encryption. In this code snippet, the IV is formed as a sequence of zeros matching the required length for the chosen ciphers IV.

The final step involves decrypting the DNA information through the utilization of the `openssl_decrypt()` function. This function requires parameters such as encrypted DNA data (`$enc_data`), encryption cipher (`$cipher`), secret key (`$secret`), encryption options (`$option`), and IV (`$iv`). The resultant decrypted DNA sequence is then preserved in `$original_plaintext` variable for further examination or processing.

In essence, this methodology guarantees that confidential DNA data stored in cloud repositories undergo secure encryption using AES-256-CBC - a well-established and sturdy encryption standard. By adhering to this systematic process, we uphold genetic information's confidentiality by shielding it from unauthorized entry while ensuring its integrity during its tenure within cloud infrastructure.

4. Result:

Table 1 includes a concise description of each test case scenario. It provides context and outlines the expected outcome for clarity. For instance, the first scenario involves a user trying to register with a username or mobile number that is already in use. The expected result is an error message indicating the username/mobile number is unavailable. This description effectively communicates the purpose of each test case. Similarly, every test case has a brief description

along with the expected results. These descriptions act as a guide for the testing team, helping them understand the scope and anticipated outcome of each test case. This approach ensures focused testing that covers all critical scenarios, guaranteeing proper functionality of the system or application.

Table 1: Test cases affecting the workflow of the web application

Test Case (TC)	Description	Expected Result	Actual Result	status
TC1	User registering through already taken User name or Mobile number	Throw an error message stating that the user name/ Mobile number is already taken	Throws an error message	Pass
TC2	The user trying to log in through an invalid username or password	Throws an error stating that the entered user credentials are correct Actual Result	Throws an error message	Pass
TC3	User entering invalid DNA Sequence	Throws an error stating that the user must enter valid data	Throws an error message	Pass
TC4	User entering in- valid Mobile Number	Throws an error stating that the user must enter valid data	Throws an error message	Pass
TC5	The user enters valid input	Successful entry	Successfully entered	Pass
TC6	Users using the system when there is no internet connectivity	Throws connection is possible due to lack of internet	Throws an error message	Pass
TC7	The user adds the data to the database connected to the cloud	Successful entry into cloud database Data added successfully	Data added successfully	Pass
TC8	The user adds data in lower or camelcase	Throws warning to change to upper case	No alert appears	Fail
TC9	The user deletes the entered data from the database	Shows alert that data is deleted successfully	Alert appears to confirm the deletion	pass
TC 10	The User adds non-existent Personal Details	Shows warning to fill in appropriate details	No alert appears	Fail
TC 11	The user already exists but tries to register again	Shows warning that account already exists	Warning Appears	Pass

5. Conclusion:

The work "Analysis and Encryption of Human Genome with Cloud Storage" has effectively tackled the crucial task of safeguarding sensitive human DNA data through advanced encryption methods. The comprehensive report shed light on the utilization of computational techniques to encrypt and securely store DNA sequences in cloud storage, particularly emphasizing the pivotal role of DNA encryption in forensic contexts. During the implementation stage, concrete results were attained, such as the creation of a fully operational website enabling users to effortlessly register, upload, access, and remove DNA files. This proficient handling of DNA sequence inputs and smooth integration with the cloud database underscores the project's technical proficiency and successful execution.

References:

1. N. Nair, T. Jain and M. Gada, "Secured File Storage in Cloud Computing Application: Secura-Drive," 2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT), Kharagpur, India, 2021, pp. 1-4, doi: 10.1109/ICCCNT51525.2021.9580145.
2. S. C. Sukumaran and M. Mohammed, "DNA Cryptography for Secure Data Storage in Cloud," Int. J. Netw. Secur., vol. 20, no. 3, pp.

447-454, May 2018. DOI: 10.6633/IJNS.201805.20(3).06.

3. D. A. Zebari and H. Haron, "Multi-Level of DNA Encryption Technique Based on DNA Arithmetic and Biological Operations," in Proc. 2018 Int. Conf. Adv. Sci. Eng. (ICOASE), Kurdistan Region, Iraq, 2018, pp. 1-4.
4. H. Gunasekaran, K. Ramalakshmi, A. Rex Macedo Arokiaraj, S. Deepa Kanmani, Chandran Venkatesan, and C. Suresh Gnana Dhas, "Analysis of DNA Sequence Classification Using CNN and Hybrid Models," Comput. Math. Methods Med., vol. 2021, Article ID 1835056, pp. 1-12, 2021, doi: 10.1155/2021/1835056.
5. S. Sudersan, M. Thangavel, B. Sowmiya, V.S. Abhijith, and P. Varalakshmi, "Enhanced DNA Cryptosystem for Secure Cloud Data Storage," in Proc. 2021 2nd Int. Conf. Secure Cyber Comput. Commun. (ICSCCC), 2021, pp. 1-4, doi: 10.1109/ICSCCC51823.2021.9478177.
6. P. V. Maitri and A. Verma, "Secure file storage in cloud computing using hybrid cryptography algorithm," in 2016 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET). IEEE, 2016, pp. 1635–1638.
7. Z. Rui and Z. Yan, “A survey on biometric authentication: Toward secure and privacy-preserving identification,” IEEE access, vol. 7, pp. 5994–6009, 2018.
8. M. Sabry, M. Hashem, T. Nazmy, and M. E. Khalifa, “Design of dna-based advanced encryption standard (aes),” in 2015 IEEE Seventh International Conference on Intelligent Computing and Information Systems (ICICIS). IEEE, 2015, pp. 390–397.
9. M. Manek, A. Chhadwa, K. Shah, M. Potey, and M. Khan, "Cloud Oriented Distributed and Encrypted File Storage (CODE-FS)," in 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA), 2018.