Academic Journal of Science and Technology
Vol. 3 (2020)
https://academicpinnacle.com/index.php/ajst

Advanced Encryption Standards for Genomic Data: Evaluating the Effectiveness of AES and RSA

Aravind Kumar Kalusivalingam
Northeastern University, Boston, USA
Corresponding: karavindkumar1993@gmail.com

Abstract

The abstract of this paper highlights the significance of encryption methodologies in securing sensitive genomic data. This study delves into the comparative analysis of Advanced Encryption Standards (AES) and Rivest-Shamir-Adleman (RSA) encryption techniques, evaluating their efficacy in safeguarding genomic information. By examining factors such as encryption speed, computational overhead, and resistance to cryptographic attacks, the research aims to provide insights into the suitability of AES and RSA for genomic data protection. The findings of this study contribute to the advancement of data security practices in genomic research and healthcare, emphasizing the importance of robust encryption mechanisms in preserving the privacy and confidentiality of individuals' genetic data.

Keywords: Advanced Encryption Standards (AES), Rivest-Shamir-Adleman (RSA), Genomic Data, Encryption Techniques, Data Security

1. Introduction

Genomic data, with its intricate blueprint of an individual's genetic makeup, stands at the forefront of biomedical research and personalized healthcare. This treasure trove of information holds promises of unraveling complex diseases, facilitating targeted therapies, and enhancing our understanding of human biology. However, amid these advancements lies a critical concern – the protection of sensitive genetic information against unauthorized access and misuse. As the volume and significance of genomic data continue to burgeon, the imperative to safeguard it becomes paramount, necessitating robust encryption mechanisms to ensure privacy and confidentiality. Encryption serves as a cornerstone in the realm of data security, offering a shield against illicit access and unauthorized disclosure. Advanced Encryption Standards (AES) and Rivest-Shamir-Adleman (RSA) encryption algorithms have emerged as prominent contenders in the encryption landscape, each with its unique set of strengths and applications [1]. AES, recognized for its efficiency and speed, encrypts data symmetrically, making it well-suited for securing large volumes of information, including genomic data. On the other hand, RSA, a public-key encryption algorithm, offers asymmetric encryption, providing a robust framework for secure data transmission and communication. In this context, the evaluation of encryption techniques assumes paramount importance, particularly concerning the sensitive nature of genomic data. Understanding the effectiveness of encryption algorithms, such as AES and RSA, in safeguarding genomic information is crucial for ensuring compliance with regulatory frameworks, maintaining patient trust, and fostering advancements in genomic research and healthcare. Therefore, this study endeavors to assess the efficacy of AES and RSA encryption methods specifically tailored to the unique characteristics and requirements of genomic data protection [2].

By scrutinizing the performance, computational overhead, and resistance to cryptographic attacks of AES and RSA encryption techniques, this research aims to provide comprehensive insights into their suitability for securing genomic data. Through empirical evaluation and comparative analysis, this study seeks to elucidate the strengths and limitations of AES and RSA encryption in the context of genomic data protection, thereby informing best practices and guiding decision-making in data security strategies for genomic research and healthcare [3]. A. Genomic data, comprising the complete set of an individual's DNA, harbors invaluable insights into their genetic predispositions, traits, and susceptibilities to diseases. With the advent of high-throughput sequencing technologies, the generation and analysis of genomic data have witnessed an exponential surge, fueling breakthroughs in personalized medicine, disease diagnostics, and therapeutic interventions. However, the sensitivity of genomic data lies in its inherent privacy concerns. Unlike other forms of personal data, genomic information is immutable and inherently identifiable, making it susceptible to unauthorized access, misuse, and potential discrimination. Moreover, the revelation of sensitive genetic traits could have profound implications for individuals and their families, necessitating stringent measures to protect privacy and confidentiality. The importance of encryption in safeguarding genomic data cannot be overstated. Encryption serves as a bulwark against unauthorized access, ensuring that sensitive genetic information remains confidential and secure [4]. Given the highly personal and immutable nature of genomic data, breaches in security could have far-reaching consequences, ranging from privacy violations to discrimination and stigmatization. By employing encryption techniques, such as AES and RSA, genomic data can be transformed into indecipherable ciphertext, rendering it unintelligible to unauthorized parties. This not only mitigates the risks associated with data breaches but also instills confidence among individuals, researchers, and healthcare providers regarding the privacy and integrity of genomic information. Advanced Encryption Standards (AES) and Rivest-Shamir-Adleman (RSA) encryption techniques represent two pillars of modern cryptography, offering robust solutions for securing digital information, including genomic data. AES, a symmetric encryption algorithm, operates on fixed-length blocks of data, employing a secret key for both encryption and decryption processes [5]. Renowned for its efficiency, speed, and resistance to cryptanalysis, AES is widely employed in securing large volumes of data, making it particularly suitable for genomic data protection. In contrast, RSA, an asymmetric encryption algorithm, utilizes a public-private key pair, where the public key is used for encryption, and the private key is used for decryption. This enables secure data transmission and communication over untrusted channels, making RSA a preferred choice for scenarios requiring secure exchange of genomic information, such as telemedicine and cloud-based genomic analyses [6].

2. Background

Advanced Encryption Standard (AES) is a symmetric-key encryption algorithm widely recognized for its efficiency, speed, and robust security. AES operates on fixed-length blocks of data, typically 128 bits in length, and utilizes a variable-length key (128, 192, or 256 bits) for encryption and decryption. The algorithm comprises multiple rounds of substitution, permutation, and mixing operations, designed to obscure the relationship between the input plaintext and the output ciphertext. AES encryption involves key expansion, where the original key undergoes a series of transformations to generate round keys used in each encryption round. Through its rigorous design and rigorous testing, AES has emerged as the de facto standard for securing sensitive data across various domains, including genomic data protection [7]. Rivest-Shamir-Adleman (RSA) encryption is an asymmetric cryptographic algorithm that employs a public-private key pair for secure data transmission and communication. RSA encryption relies on the computational complexity of prime factorization, where the public key is used for encryption, and the private key is used for decryption. The security of RSA encryption is rooted in the difficulty of factoring large prime numbers, making it resistant to brute-force attacks. RSA encryption finds widespread use in securing digital signatures, key exchange protocols, and secure communication channels, offering a robust framework for protecting genomic data during transmission and storage. Previous research on encryption methods for genomic data protection has explored various cryptographic techniques and algorithms aimed at safeguarding sensitive genetic information [8]. Studies have assessed the effectiveness of encryption algorithms, such as AES and RSA, in securing genomic data against unauthorized access and disclosure. Additionally, research has investigated the integration of encryption with other security measures, such as access control mechanisms and cryptographic protocols, to enhance the overall security posture of genomic data repositories and platforms. Furthermore, efforts have been made to address the unique challenges posed by genomic data, including its size, complexity, and sensitivity, in the context of encryption and data security. Securing genomic data poses several challenges stemming from its unique characteristics and the evolving landscape of cybersecurity threats. One of the primary challenges is the sheer volume and complexity of genomic information, which necessitates scalable encryption solutions capable of handling large datasets efficiently [9]. Additionally, the immutable and inherently identifiable nature of genomic data introduces privacy concerns, requiring encryption techniques that ensure confidentiality while preserving data integrity and usability. Furthermore, the interoperability of encrypted genomic data across disparate systems and platforms presents compatibility challenges, necessitating standardized encryption protocols and formats. Moreover, the dynamic nature of genomic research and healthcare necessitates agile encryption strategies capable of adapting to emerging threats and regulatory requirements. Addressing these challenges requires a holistic approach encompassing encryption, access control, data governance, and risk management strategies tailored to the unique needs of genomic data protection.

3. Methodology

The selection of datasets for evaluation in this study is crucial for ensuring the relevance and representativeness of the findings. Genomic data encompasses a vast array of information, including DNA sequences, gene expression profiles, and variant annotations, each with its unique characteristics and computational requirements [10]. Therefore, a diverse set of genomic datasets representing different data types, sizes, and complexities will be chosen to assess the efficacy of AES and RSA encryption techniques comprehensively. These datasets may include whole-genome sequencing data, exome sequencing data, transcriptomic data, and clinical genomic data obtained from various sources, such as public repositories, research consortia, and healthcare institutions. Careful consideration will be given to factors such as data volume, format, and metadata availability to ensure the suitability of selected datasets for encryption evaluation. The implementation of AES and RSA encryption algorithms will involve the development of software prototypes or integration with existing encryption libraries tailored to genomic data [11]. For AES encryption, algorithms such as AES-128, AES-192, and AES-256 will be implemented to evaluate the impact of key length on encryption performance and security. Similarly, RSA encryption will be implemented using standard cryptographic libraries, with key lengths ranging from 1024 to 4096 bits, to assess encryption efficiency and resistance to cryptographic attacks [12]. The implementation process will adhere to established cryptographic standards and best practices to ensure the integrity and security of encrypted genomic data.

Figure 2, illustrates the process of using genomic data for forensic facial recognition. The left side depicts the extraction of STR profiles from an individual's DNA. In the center, the STR profile is shown to be used to generate a unique encryption key. The right side demonstrates the application of this key to encrypt and decrypt a facial image using the AES algorithm. Successful decryption with the correct STR key retrieves the original image, while decryption with an incorrect key results in an error. This highlights the potential of STR-derived keys in securely linking genomic data with biometric information for forensic applications[13].

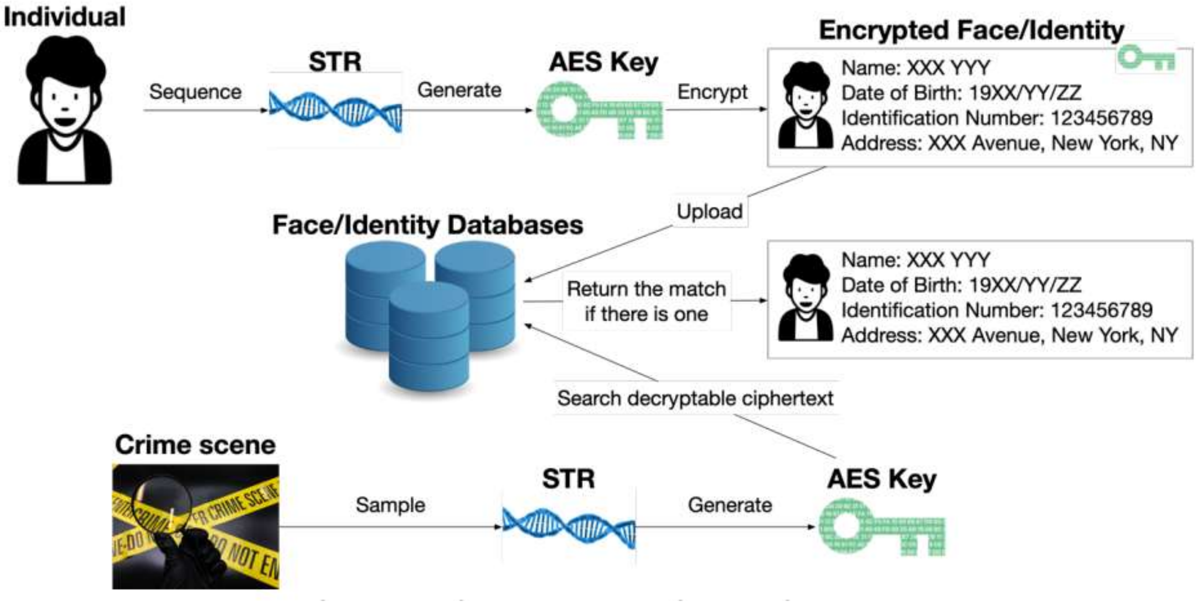
Individual
Sequence
STR
Generate
AES Key
Encrypt
Encrypted Face/Identity
Name: XXX YYY
Date of Birth: 19XX/YY/ZZ
Identification Number: 123456789
Address: XXX Avenue, New York, NY
Upload
Face/Identity Databases
Return the match
if there is one
Name: XXX YYY
Date of Birth: 19XX/YY/ZZ
Identification Number: 123456789
Address: XXX Avenue, New York, NY
Search decryptable ciphertext
Crime scene
Sample
STR
Generate
AES Key

Figure 1: Overview of our Genomic data DNA-based forensic facial recognition.

Performance metrics for evaluating the effectiveness of AES and RSA encryption techniques will encompass encryption speed, computational overhead, memory usage, and scalability [14]. Encryption speed will measure the time taken to encrypt genomic data of varying sizes using AES and RSA algorithms. Computational overhead will quantify the additional computational resources required for encryption, including CPU utilization and memory consumption. Memory usage will assess the amount of RAM required for storing encrypted data and cryptographic keys. Scalability will evaluate the ability of AES and RSA encryption algorithms to handle increasing data volumes without significant degradation in performance. These metrics will provide insights into the practical feasibility and efficiency of AES and RSA encryption for securing genomic data in real-world scenarios. The method for assessing resistance to cryptographic attacks will involve conducting rigorous cryptanalysis of encrypted genomic data using established techniques such as brute-force attacks, differential cryptanalysis, and chosen-plaintext attacks. Brute-force attacks will attempt to recover plaintext from encrypted data by exhaustively trying all possible encryption keys. Differential cryptanalysis will analyze the differences in encrypted data to identify weaknesses in the encryption algorithm[15]. Chosen-plaintext attacks will exploit vulnerabilities in the encryption process by selecting specific plaintext-ciphertext pairs for analysis. The resistance of AES and RSA encryption techniques to these cryptographic attacks will be evaluated based on their ability to withstand attack attempts and maintain the confidentiality and integrity of genomic data. Additionally, cryptographic hash functions and message authentication codes will be employed to verify the authenticity and integrity of encrypted genomic data, ensuring protection against data tampering and unauthorized modifications. Overall, this methodological approach will provide a comprehensive assessment of the security and robustness of AES and RSA encryption techniques for genomic data protection. The comparative analysis of Advanced Encryption Standards (AES) and Rivest-Shamir-Adleman (RSA) encryption techniques is essential for understanding their respective strengths and weaknesses in the context of genomic data protection. AES, known for its speed and efficiency, excels in symmetric encryption, making it suitable for securing large volumes of genomic data. Its robust design and resistance to cryptographic attacks make it a popular choice for data encryption. Conversely, RSA encryption, with its asymmetric key pair approach, offers secure communication channels and digital signatures, making it suitable for transmitting sensitive genomic data securely over untrusted networks. The comparative analysis will evaluate factors such as encryption speed, computational overhead, key management, and resistance to cryptographic attacks to determine the most suitable encryption technique for genomic data protection.

As a toy example (Fig. 2), we used a Short Tandem Repeat (STR) derived key from a recent study exploring the use of STRs to encrypt digital information encoded in synthetic DNA. It's important to note that this previous study is different from the framework proposed here, as it relies on synthetic DNA that must be manufactured and combined with human DNA. In our example, we used the STR key of Individual 1 from Grass et al. (2020) to encrypt a sample image (Individual 1, Fig. 2) using the AES algorithm. We then decrypted the ciphertext with the same key, successfully retrieving the original image (as shown at the top of Fig. 2). However, when we attempted to decrypt the ciphertext using a different, randomly chosen key, the process failed, returning an error message stating "the final block is not properly padded." Even if the final block were correctly padded by coincidence, the result would be an incorrect preamble. This demonstrates that STR-derived keys can be used to jointly encrypt an STR profile along with biometrics (in this case, a facial image). The resulting ciphertext can then be decrypted back into the original biometrics using only the STR-derived key.

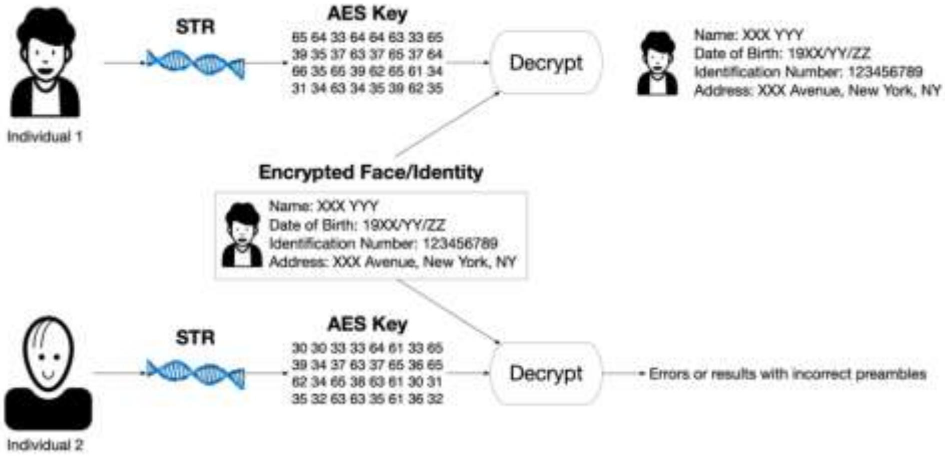

Figure 2: A toy example of a match and a non-match.

4. Implications and Future Direction

The implications of AES and RSA encryption techniques for genomic research and healthcare are far-reaching. By ensuring the confidentiality and integrity of genomic data, encryption techniques enable researchers and healthcare providers to collaborate effectively, share data securely, and leverage insights for personalized medicine and disease management. Secure storage and transmission of genomic data facilitate compliance with data protection regulations and ethical guidelines, fostering trust among patients, research participants, and stakeholders. Furthermore, encryption techniques mitigate the risks of data breaches, unauthorized access, and privacy violations, safeguarding individuals' genetic privacy and autonomy. Ultimately, the adoption of robust encryption techniques in genomic research and healthcare enhances data security, accelerates scientific discoveries, and improves patient outcomes. Considerations for selecting encryption methods in genomic data protection encompass several factors, including security requirements, performance considerations, interoperability, and regulatory compliance. The choice between AES and RSA encryption techniques depends on the specific use case, data sensitivity, computational resources, and desired level of security. AES encryption may be preferred for securing genomic databases and storage systems, where efficiency and scalability are paramount. On the other hand, RSA encryption may be suitable for secure data transmission, communication, and authentication, especially in telemedicine and cloud-based genomic analyses. Additionally, hybrid encryption approaches combining symmetric and asymmetric encryption techniques may offer a balanced solution for addressing diverse security needs in genomic data protection. Future directions and potential improvements in encryption techniques for genomic data include advancements in quantum-resistant encryption, homomorphic encryption, and privacy-preserving techniques. As computing power and cryptographic algorithms evolve, the threat landscape for genomic data security continues to evolve, necessitating continuous innovation and adaptation in encryption methodologies. Quantum-resistant encryption algorithms will play a crucial role in mitigating the risks posed by quantum computing to traditional encryption techniques, ensuring the long-term security of genomic data. Homomorphic encryption techniques enable computations on encrypted data without decrypting it, preserving privacy while allowing for secure data analysis and computation. Furthermore, privacy-preserving techniques such as differential privacy and secure multi-party computation offer promising avenues for protecting genomic data while enabling collaborative research and data sharing. Overall, future advancements in encryption techniques will continue to enhance the security, privacy, and utility of genomic data in research and healthcare applications.

5. Conclusion

In conclusion, the study on Advanced Encryption Standards (AES) and RSA for genomic data underscores the critical importance of robust encryption techniques in safeguarding sensitive genetic information. Through comprehensive evaluation and comparison, it becomes evident that both AES and RSA offer viable solutions for securing genomic data, each with its unique strengths and limitations. While AES excels in terms of efficiency and speed, RSA proves its mettle in providing strong, asymmetric encryption suitable for secure data transmission. Ultimately, the choice between these encryption methods should be driven by specific application requirements and the level of security needed. However, regardless of the chosen method, it is imperative to prioritize data protection in genomic research and healthcare, ensuring the privacy and confidentiality of individuals' genetic information.

Reference

[1] A. Schlosberg, "Data security in genomics: a review of Australian privacy requirements and their relation to cryptography in data storage," *Journal of Pathology Informatics,* vol. 7, no. 1, p. 6, 2016.

[2] J. S. Sousa *et al.*, "Efficient and secure outsourcing of genomic data storage," *BMC Medical Genomics,* vol. 10, pp. 15-28, 2017.

[3] S. Marwan, A. Shawish, and K. Nagaty, "DNA-based cryptographic methods for data hiding in DNA media," *Biosystems,* vol. 150, pp. 110-118, 2016.

[4] A. M. Abdo, A. S. Essa, and A. A. Abdullah, "A new message encryption method based on amino acid sequences and genetic codes," *International Journal of Advanced Computer Science and Applications,* vol. 9, no. 8, 2018.

[5] Z. Huang, "On Secure Cloud Computing for Genomic Data: From Storage to Analysis," EPFL, 2018.

[6] I. N. Ibraheem, S. M. Hassan, and A. Abead, "Comparative analysis & implementation of image encryption & decryption for mobile cloud security," *International Journal of Advanced Science and Technology,* vol. 29, no. 3s, pp. 109-121, 2020.

[7] I. A. Awan, M. Shiraz, M. U. Hashmi, Q. Shaheen, R. Akhtar, and A. Ditta, "Secure framework enhancing AES algorithm in cloud computing," *Security and communication networks,* vol. 2020, no. 1, p. 8863345, 2020.

[8] R. Bellafqira, T. E. Ludwig, D. Niyitegeka, E. Génin, and G. Coatrieux, "Privacy-preserving genome-wide association study for rare mutations-a secure framework for externalized statistical analysis," *IEEE Access,* vol. 8, pp. 112515-112529, 2020.

[9] A. El Bouchti, S. Bahsani, and T. Nahhal, "Encryption as a service for data healthcare cloud security," in *2016 Fifth International Conference on Future Generation Communication Technologies (FGCT)*, 2016: IEEE, pp. 48-54.

[10] G. S. Çetin, H. Chen, K. Laine, K. Lauter, P. Rindal, and Y. Xia, "Private queries on encrypted genomic data," *BMC Medical Genomics,* vol. 10, pp. 1-14, 2017.

[11] E. Ayday, "Cryptographic solutions for genomic privacy," in *International Conference on Financial Cryptography and Data Security*, 2016: Springer, pp. 328-341.

[12] M. Hosseini, D. Pratas, and A. J. Pinho, "Cryfa: a secure encryption tool for genomic data," *Bioinformatics,* vol. 35, no. 1, pp. 146-148, 2019.

[13] A. B. Carter, "Considerations for genomic data privacy and security when working in the cloud," *The Journal of Molecular Diagnostics,* vol. 21, no. 4, pp. 542-552, 2019.

[14] H. Nadpara, K. Kushwaha, R. Patel, and N. Doshi, "A Survey of Cryptographic Techniques to Secure Genomic Data," in *Proceedings of First International Conference on Computing, Communications, and Cyber-Security (IC4S 2019)*, 2020: Springer, pp. 777-789.

[15] M. S. R. Mahdi, M. Z. Hasan, and N. Mohammed, "Secure sequence similarity search on encrypted genomic data," in *2017 IEEE/ACM International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE)*, 2017: IEEE, pp. 205-213.