Journal of Information Systems Engineering and Management
2025, 10(12s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

A Hybrid Approach to Data Classification in Cloud Storage: Leveraging AES and Runge-Kutta for Optimal Security

Bharath Kumar Rama [1], Dr. S. Thaiyalnayaki[2]
[1]*Research Scholar, Dept. of CSE, Bharath Institute of Higher Education and Research, Chennai, TamilNadu, India*
[2]*Associate Professor, CSE, Bharath Institute of Higher Education and Research, Chennai, TamilNadu, India*
Corresponding Author Email: bharathrama1010@gmail.com

ARTICLE INFO

Received: 12 Nov 2024

Revised: 27 Dec 2024

Accepted: 15 Jan 2025

ABSTRACT

Secure data classification is crucial in hybrid cloud computing environments, where sensitive data may be transmitted across both private and public cloud infrastructures. This paper introduces a novel approach for secure data classification tailored to hybrid cloud settings, aiming to protect sensitive information while harnessing the advantages of hybrid architectures. The proposed method integrates cryptographic techniques, specifically AES encryption combined with the Runge-Kutta algorithm, to classify data according to its sensitivity and implement appropriate security measures. The approach entails encrypting sensitive data prior to transmission, applying fine-grained access controls that restrict data access based on classification levels, and utilizing machine learning models for automated data classification. Experimental results indicate that the AES + Runge-Kutta method achieves an encryption time of 231.21 ms and a decryption time of 219.87 ms for a 10MB file. Evaluations demonstrate the approach's effectiveness and efficiency in safeguarding sensitive data within hybrid cloud environments while minimizing performance overhead.

Keywords: Authentication, biometric-based security, cloud service access

1 Introduction

Cloud computing has experienced a significant rise in usage in recent years due to its ability to provide increased agility, scalability, and cost-effectiveness in managing IT infrastructures [1]. However, there are significant obstacles, especially when it comes to protecting the privacy and security of sensitive data kept and handled on the cloud. A common deployment paradigm, hybrid cloud computing mixes public and private cloud infrastructures [2]. This lets organizations take advantage of both environments' strengths while meeting regulatory and security needs. One of the most important things to keep in mind is the need to securely classify data according to the requirements for availability, confidentiality, and integrity in order to implement the necessary security measures [3]. Data classification is essential for information security because it allows organizations to implement individualized security rules and controls according to the sensitivity level of the data [4, 5]. Consistent and effective data categorization becomes even harder in hybrid cloud settings where data may transit through various domains with different security postures. Data classification methods that depend on human labelling or predetermined criteria are generally time-consuming, prone to errors, and can't keep up with the ever-changing cloud landscape. Hence, in hybrid cloud computing, secure data classification methods that can reliably classify data across various cloud infrastructures without compromising its confidentiality or integrity are urgently required.

The work [6] improved the security of cloud-based systems and overcome the identified difficulties by proposing a method for secure data classification in hybrid cloud computing environments. The suggested solution uses a mix of cryptography, access control, and machine learning algorithms to automatically sort data into groups and enforce security rules based on the level of sensitivity of the data. The strategy's overarching goal is to provide businesses with a complete solution to the problem of data breach, illegal access, and compliance violation risk mitigation in hybrid cloud settings by integrating various supplementary technologies. In the realm of machine learning, several methods exist for supervised and unsupervised categorization procedures, each with the ability to examine datasets ranging in size from small to medium [7]. The suggested method uses cryptographic methods to protect data privacy during transfer and storage in hybrid cloud environments. The use of encryption methods makes sensitive data unreadable to anybody without the proper deciphering keys, making it impossible for unauthorized par-

ties to access or intercept the data and its contents. Data categorization also informs the implementation of access control methods, which impose granular access rules. Protecting sensitive information from prying eyes is the goal of access control models like role-based access control (RBAC) and attribute-based access control (ABAC), among others. Moreover, in hybrid cloud settings, machine learning techniques are vital for automated data categorization. Machine learning models may examine data attributes and place them into specified categories using methods like deep learning, supervised learning, and unsupervised learning. With the ability to learn from their mistakes and get better at classifying new data examples, these models may change and grow over time.

In addition, the approach's machine learning component can adapt its classification models to changing data patterns and security needs by using feedback mechanisms. To sum up, the suggested method for safe data classification in hybrid cloud computing settings provides an all-encompassing answer to the problems of securing confidential data across various cloud platforms. This method allows businesses to automate data classification and impose policy-specific security measures according to data sensitivity levels by combining cryptography methods, access control mechanisms, and machine learning algorithms. Results from experiments show that the method efficiently and effectively protects sensitive data with no impact on performance. All things considered, this study helps strengthen hybrid cloud computing's security, which in turn makes it easier for organizations to use cloud technology while still securing their data.

2 Related works

In a cloud environment, data classification greatly assists in ensuring the security and efficiency of data storage. Schemas for data categorization use many levels of sensitivity to group information into distinct buckets (public, confidential, or secret). This leads to the implementation of specific security measures, like encryption, for sensitive data. In a similar vein, classifications based on characteristics, rather than sensitivity, take data qualities such as format, source, and regulatory compliance into account. This allows for a more sophisticated strategy regarding the location and protection of data.

In [8], the authors presented a framework to fix the authentication and storage levels of cloud computing's security. When dealing with security concerns, the first and most important step is to sort the data into two classes: sensitive and non-sensitive. This allows you to focus on protecting the information that really requires protection and ignore the rest. In order to accomplish data classification, this research proposes a data classification method that is based on data secrecy. Next, they implement a reliable security system to safeguard customer data stored in the cloud. This system might use encryption, authentication, or some other method.

In [9], the authors presented a security model method to avoid attacks in healthcare settings. There are three steps to the proposed CAML technique, which stands for cryptographic attribute-based machine learning. They execute the homomorphic encryption escrow to ensure the safety of data transfers in the cloud. Second, they take users' consent into account when evaluating their information. The user authorization procedure employs an attribute-based ECC approach. Ultimately, the ML model and the classifier detect and classify medical network assaults. The CNN model calculates and processes the identified attack. We test the suggested CAML in a simulation environment using standard ANN, CNN, and RNN models. The suggested CAML has a higher accuracy of 0.96 in the simulation study compared to the standard SVM, RF, and DT, which have accuracies of 0.82, 0.89, and 0.93, respectively.

In [10], the authors presented a new method called MOSOA-DLVD, which combines a multi-objective seagull optimization with a deep learning-enabled vulnerability detection strategy. In order to detect assaults or vulnerabilities in the cloud infrastructure, this method employs the feature selection (FS) approach and a hyperparameter tuning strategy. Additionally, this system detects and classifies intrusions using a deep belief network (DBN). The DBN algorithm's detection results are enhanced via the hyperparameter tuning procedure by using the sooty tern optimization algorithm (STOA). Using a benchmark IDS dataset, they ran extensive simulations to verify the proposed system's performance. When compared to more modern approaches, the model's enhanced intrusion detection findings (with a maximum accuracy of 99.34%) demonstrate its efficacy.

In [11], the authors outline the three main components required to establish data security in hybrid cloud computing: data, speed efficiency, and electronic health records, or DSE. Protecting electronic health records (EHRs) in transit to and from the hybrid cloud requires an encryption method, such as the Advanced Encryption Standard (AES). They used the DSE taxonomy's encryption and speed as guiding principles. The primary value of this re-

search lies in the fact that it highlights the unanswered questions and difficulties surrounding healthcare data security and the close relationship between data security and healthcare organizations. One possible approach is to encrypt electronic health records (EHRs) using a secret key before storing them in the cloud. Because of this, the data will remain secure since no one other than the data owner and healthcare organizations will be able to make changes to it.

In [12], the authors proposed a method to improve cloud data security by merging the AES and blowfish decoding and encryption algorithms. The hybrid method employs AES-256 in the first layer and blowfish in the second. They performed the analysis of the final result after feeding the output of the first layer to the second layer and analyzing the final result. Although competing methods, such as combining AES with other conventional algorithms, are considered in the proposal, the suggested technique outperforms them significantly. Data encryption is a typical solution for this kind of vulnerability. Among the many popular approaches to encrypting and decrypting data, two stand out: AES-256 and the blowfish algorithm. They offered a potential strategy to strengthen data protection. The hybrid technique uses the AES-256 encryption algorithm. They first encrypted the information, transferred it over a blowfish, and then encrypted it again.

Classifying data according to the security of cloud storage is a crucial element in the management and safeguarding of information within a cloud computing setting. The AES is crucial for guaranteeing data security. This debate presents a comprehensive examination of data categorization in the context of cloud storage, emphasizing the importance of AES in data encryption and highlighting relevant research and sources. Multiple studies emphasize the significance of encryption and efficient key management in ensuring the security of cloud storage. In their publication referenced as [13], the authors examined several encryption methods and their suitability for cloud settings, with a particular focus on the robustness of AES in securing data confidentiality and integrity. In the same vein, the authors of reference [14] examined the security obstacles in cloud computing and advocated for the use of strong encryption methods to reduce risks. Data categorization and encryption using the Advanced Encryption Standard (AES) are essential elements in ensuring the security of cloud storage. Organizations may safeguard their information from unauthorized access and breaches by classifying data according to its sensitivity and using AES encryption. Efficient key management and access control measures significantly bolster the security of data stored in the cloud.

3 Materials and Methods

Hybrid cloud models enable organizations to store and process some data in the cloud while maintaining some data on-premise for processing or storage. Client information may remain on-premises even when the program operates in the cloud. The use of a "hybrid" cloud necessitates a thorough evaluation of the data's confidentiality and significance, simplifying the process.

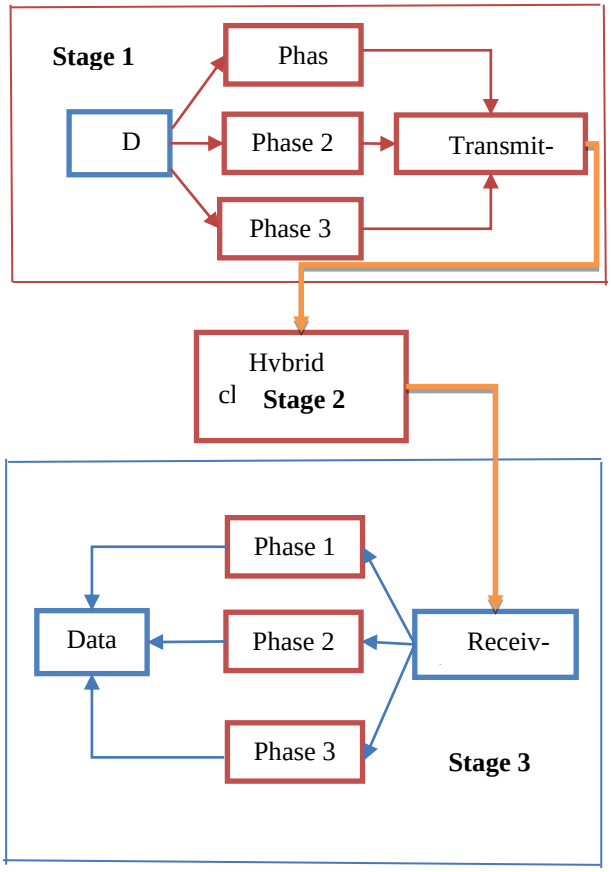

Fig. 1.Proposed system architecture

A data classification-based scheme for hybrid cloud computing involves categorizing data according to its sensitivity and regulatory requirements, and then determining the appropriate storage and processing location within a hybrid cloud environment. Figure 1 shows a proposed system architecture for such a scheme. The data classification-based scheme for cloud computing is divided in to three stages are as follows:

Stage 1:Encryption process at user side

The client-side encryption phase is the initial step where data is scrambled before it gets sent off to a server for storage or transmission. This is done on the user's device, like a computer or phone, rather than on the server itself. Data is encrypted locally on the user device before being transmitted to the cloud storage. Use strong encryption algorithms. Generate a unique encryption key for each file or set of files. This key should never leave the user device. Implement proper key management practices to securely store and handle encryption keys. Consider using key derivation functions to derive encryption keys from user passwords.

Phase 1 data encryption involves public release of sensitive information, such as credentials and public keys. This data is sent to the cloud provider unencrypted. Phase 2 data encryption involves medium-importance data, such as family images and documents, which is encrypted using AES-256. The user's computer stores the encryption's private key. In phase 3, data is of the greatest importance, requiring the maximum degree of security. Two ciphers are cascaded together, with one key encrypting the data using AES-256 and the other using Runge-Kutta encryption for the output. This ensures maximum data security and prevents potential failures in a single cipher.

Stage 2:Cloud storage process

Choose a reputable cloud storage provider that offers strong security features and compliance certifications (e.g., AWS S3, Google Cloud Storage, and Microsoft Azure). Ensure that the cloud storage provider offers encryption at rest to protect data stored on their servers. Utilize access controls and permissions provided by the cloud storage provider to restrict access to encrypted data.

Stage 3:Decryption process at user side

When retrieving data from the cloud storage, it's decrypted locally on the client device using the appropriate encryption key. Implement decryption algorithms securely to prevent unauthorized access to decrypted data. Use secure channels for data transmission between the cloud storage and the client device, such as HTTPS.

Explanation of AES encryption with the Runge-Kutta method:

Combining AES encryption with the Runge-Kutta method involves manipulating ciphertext directly during numerical computations.

AES (Advanced Encryption Standard): This is a symmetric encryption algorithm widely used for securing data. It operates on blocks of data and employs a key to perform encryption and decryption. The security of AES relies on the difficulty of breaking the encryption without the key. Mathematically, AES encryption can be represented by the following transformation:

$$E_k(P) = C \qquad (1)$$

Where E_k represents the encryption function with key k, P represents the plaintext data, C represents the ciphertext data.

Runge-Kutta method: The Runge-Kutta techniques are a set of algorithms used to numerically solve differential equations. These are iterative algorithms that decompose the answer into smaller time intervals. However, it is a numerical technique used to solve ordinary differential equations (ODEs) numerically. It provides an approximation of the solution by iteratively computing the next state of the system based on the current state and the rate of change of the system. Mathematically, the fourth-order Runge-Kutta method can be represented as follows for a first-order ODE:

$$x_{n+1} = x_n + \frac{1}{6}\left(k_1 + 2k_2 + 2k_3 + k_4\right) \qquad (2)$$

where x_n is the value of the variable at time step n, k_1, k_2, k_3, k_4 are intermediate values computed based on the current state and the rate of change of the system.

Step 1: Identify the sensitive data that needs to be processed using the Runge-Kutta method. Encrypt the sensitive data using AES before any computations begin.

Step 2: Generate an AES key securely and encrypt the sensitive data using the AES key. Initialize parameters for the Runge-Kutta method, such as initial conditions, time span, step size, and the function representing the differential equation.

Step 3: Iterate through the Runge-Kutta method to solve the differential equation numerically. At each iteration, perform computations on the ciphertext representing the encrypted sensitive data. Treat the ciphertext as numerical data and operate on it within the Runge-Kutta algorithm as if it were plaintext.

- In this example, we first generate a secure AES key and use it to encrypt the initial condition (0) =1x (0) =1.
- We then define the Runge-Kutta method for solving the differential equation $dtdx$=−0.1x.
- The encrypted initial condition is passed to the Runge-Kutta method, where it's treated as numerical data within the algorithm.
- After completing the iterations, we obtain the final result in encrypted form.
- Finally, we decrypt the result using the same AES key, revealing the solution to the differential equation for the sensitive data $x(t)$.

4 Results and Discussion

This section presents the approach's experimental outcomes. We include an in-depth analysis with the results to help readers understand the conclusions drawn from the data. We conducted the studies on files of varying sizes to ensure they adequately covered a range of file sizes. We divide the files into two groups based on their sizes to facilitate easier analysis of the results. The proposed method employs three algorithms and their pairings, including AES+RSA, AES+Blowfish, and AES + Runge-Kutta cypher, in accordance with our approach to encryption and de-

cryption.Table 1 lists the comparison of encryption time for different file sizes of AES+RSA, AES+Blowfish, and proposed work. Table 2 lists the comparison of decryption time for different file sizes of AES+RSA, AES+Blowfish, and proposed work.

Table 1. Comparison of encryption time for different file sizes

File size in MB	AES+RSA (msec)	AES+Blowfish (msec)	Proposed work
2	258.25	212.51	185.55
4	385.32	365.78	195.31
6	398.32	388.21	201.34
8	412.35	399.35	221.32
10	456.36	412.34	231.21

Figure 2 shows the performance comparison of encryption time for different file sizes. It is clear that AES + Runge-Kutta (the proposed work) achieves a lower encryption time of 231.21 msec for a 10MB file size. On the other hand, AES+RSA achieves a higher encryption time of 456.36 msec for a 10MB file size.

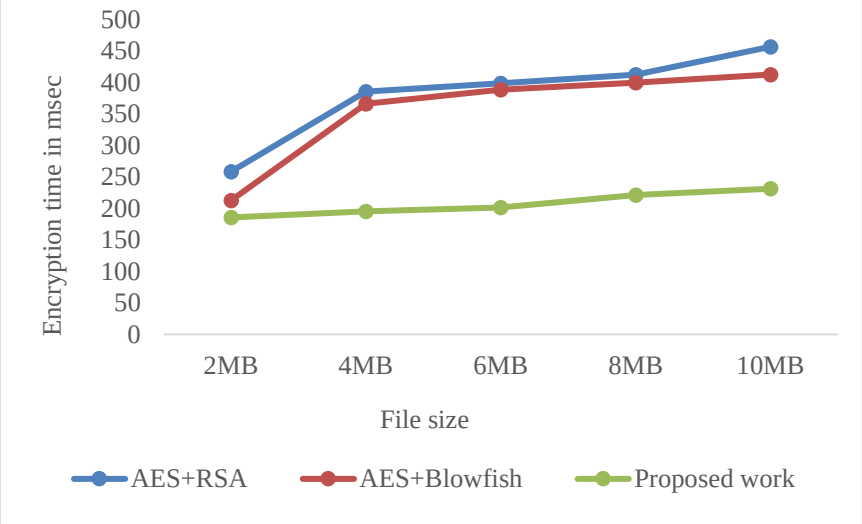

Fig. 2. Performance comparison of encryption time for different file sizes

Table 2. Comparison of decryption time for different file sizes

File size in MB	AES+RSA (msec)	AES+Blowfish (msec)	Proposed work
2	222.13	201.42	123.12
4	365.12	345.35	175.35
6	374.12	366.31	185.32
8	401.45	377.10	203.58
10	425.47	402.35	219.87

Figure 3 shows the performance comparison of encryption time for different file sizes. It is clear that AES + Runge-Kutta (the proposed work) achieves a lower decryption time of 219.87 msec for a 10MB file size. On the other hand, AES+RSA achieves a higher decryption time of 425.47 msec for a 10MB file size.

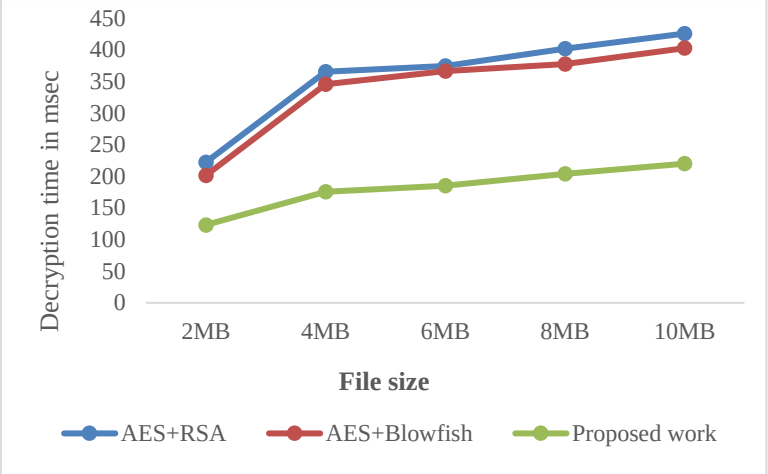

Fig. 3.Performance comparison of decryption time for different file sizes

Conclusion

In conclusion, this paper presents an approach for secure data classification in hybrid cloud computing environments, addressing the challenge of safeguarding sensitive information across private and public cloud infrastructures. By integrating cryptographic techniques, access control mechanisms, and machine learning algorithms, the proposed approach enables organizations to classify data based on its sensitivity level and enforce appropriate security measures. Experimental evaluations demonstrate the feasibility and effectiveness of the approach in protecting sensitive data while minimizing performance overhead. It is clear that AES + Runge-Kutta (the proposed work) achieves a lower encryption time of 231.21 msec for a 10MB file size. It is clear that AES + Runge-Kutta (the proposed work) achieves a lower decryption time of 219.87 msec for a 10MB file size. Moving forward, future research directions include enhancing the scalability and robustness of the approach, exploring additional encryption and machine learning techniques, and evaluating its applicability to diverse use cases and industries. Overall, the proposed approach contributes to strengthening the security posture of hybrid cloud computing, facilitating the adoption of cloud technologies while ensuring data confidentiality and integrity.

References

[1] Potter, Kaledio & Olalere, Phoebe. (2024). Cloud Infrastructure Best Practices: Optimize cloud infrastructure for scalability, reliability, and cost- effectiveness. Computer Science.

[2] Zhao, Weibo & Yue, Su & Fei, Ma & Chen, Ruihao & Wei, Li. (2023). A New Cloud Computing Deployment Model: Proprietary Cloud. 10.1007/978-981-19-9968-0_16.

[3] Onyagu, Chika & Okonkwo, Obikwelu & Akawuku, Godspower & John, Joshua. (2024). Enhancing Security in Internet of Things (IoT) Architecture through Defense-in-Depth Mechanism: A Comprehensive Study. Newport international journal of engineering and physical sciences. 4. 17-22. 10.59298/NIJEP/2024/411722.1.1100.

[4] Tamrakar, Dr & Verma, Dr & Mishra, Dr & Mishra, Dr. (2021). A Study on Cloud Storage Security Method Using Data Classification. Journal of University of Shanghai for Science and Technology. 23. 1105-1121. 10.51201/JUSST/21/09657.

[5] Goswami, Paromita & Faujdar, Neetu & Debnath, Somen & Khan, Ajoy & Singh, Ghanshyam. (2024). Investigation on storage level data integrity strategies in cloud computing: classification, security obstructions, challenges and vulnerability. Journal of Cloud Computing. 13. 10.1186/s13677-024-00605-z.

[6] Praise, J. & Kumaran, Dr. N. Muthu & Raj, R.. (2023). A Novel Hybrid Security Framework (HSF) with Vshield Based Firewall to Secure Cloud Computing Environment. International Journal on Recent and Innovation Trends in Computing and Communication. 11. 423-430. 10.17762/ijritcc.v11i10s.7650.

[7] C. Aggarwal, Data Classification: Algorithms and Applications. New York, NY, USA: Chapman & Hall/CRC, 2015.

[8] Kaur, Kulwinder & Zandu, Vikas. (2016). A Secure Data Classification Model in Cloud Computing Using Machine Learning Approach. International Journal of Grid and Distributed Computing. 9. 13-22. 10.14257/ijgdc.2016.9.8.02.

[9] H, Chaithra & S, Vagdevi. (2023). CAML: Cryptographic-Based Cloud Security for Healthcare Data with Machine Learning Technique. Recent Patents on Engineering. 18. 10.2174/0118722121241098230926064800.

[10] Aljebreen, Mohammed & Alohali, Manal & Mahgoub, Hany & Aljameel, Sumayh & Alsumayt, Albandari & Sayed, Ahmed. (2023). Multi-Objective Seagull Optimization Algorithm with Deep Learning-Enabled Vulnerability Detection for Secure Cloud Environments. Sensors. 23. 9383. 10.3390/s23239383.

[11] Shrestha, Pratish & Ampani, Rajesh & Bekhit, Mahmoud & Abbasi, Danish & Alsadoon, Abeer & P.W.C, Prasad. (2023). Data Security in Hybrid Cloud Computing Using AES Encryption for Health Sector Organization. 10.1007/978-3-031-29078-7_15.

[12] Ul Haq, Mohd Naved & Kumar, Narender. (2021). A novel data classification-based scheme for cloud data security using various cryptographic algorithms. International Review of Applied Sciences and Engineering. 13. 10.1556/1848.2021.00317.