Improved Association Rule Mining-Based Data Sanitization for Privacy Preservation Model in Cloud

Rajkumar Patil and Gottumukkala HimaBindu

Department of Computer Science Engineering, GITAM (Deemed to be) University, Hyderabad, Telangana, India

https://doi.org/10.26636/jtit.2023.166922

Abstract — Data security in cloud services is achieved by imposing a broad range of privacy settings and restrictions. However, the different security techniques used fail to eliminate the hazard of serious data leakage, information loss and other vulnerabilities. Therefore, better security policy requirements are necessary to ensure acceptable data protection levels in the cloud. The two procedures presented in this paper are intended to build a new cloud data security method. Here, sensitive data stored in big datasets is protected from abuse via the data sanitization procedure relying on an improved apriori approach to clean the data. The main objective in this case is to generate a key using an optimization technique known as Corona-integrated Archimedes Optimization with Tent Map Estimation (CIAO-TME). Such a technique deals with both restoration and sanitization of data. The problem of optimizing the data preservation ratio (IPR), the hiding ratio (HR), and the degree of modification (DOM) is formulated and researched as well.

Keywords — data security, improved apriori, modified data restoration, sanitization.

1. Introduction

Cloud computing (CC) creates a massive virtualized data resource pool with various services connecting a huge number of resources [1]–[3] to achieve the required levels of portability and reliability. The three fundamental service delivery models include the following: infrastructure as a service, platform as a service, and software as a service, and the four development trends observed in the world of CC are public cloud, private cloud, cloud platform, public clouds, and virtual private cloud [4]–[6].

IaaS is a service-managed, computer hardware and fixed network provisioning concept with significant expansion capabilities. PaaS is a service paradigm that uses the middleware of the service model to offer integrated development environments, frameworks, applications, and development tools [7]–[9], while SaaS is a category of remote computing services. All three technologies are used in cloud computing. However, because users are not aware of the sources or owners of specific resources in such a structure, it is more challenging to protect such resources and services against attacks. The task is even more complicated, because many companies owning and running community groups, as well as outside stakeholders, operate in the capacity of cloud administrators. To benefit from the advantages of multiple cloud deploy-

Tab. 1. List of abbreviations used in this paper.

Abbreviation	Description
AOA	Archimedes optimization algorithm
ARM	Association rule mining
BMO	Blue monkey optimization
BES	Bald eagle search optimization
CHIO	Corona virus herd immunity algorithm
CIAO-TME	Corona integrated Archimedes optimization with tent map estimation
CNN	Convolutional neural network
CPA	Chosen plaintext attack
CP-ABE	Cipher text policy attribute-based encryption
CHIO	Corona virus herd immunity optimization
CC	Cloud computing
DOM	Degree of modification
EHR	Electronic health record
IDS	Intrusion detection system
IPR	Information preservation ratio
HLPN	High-level Petri nets
HBA	Honey badger algorithm
HPA	Honey pot algorithm
HR	Hiding ratio
IF	Impact factor
IM	Information masking
IPR	Information preservation ratio
GWT	Gabor wavelet transform
GMGW	Genetically modified glowworm swarm optimization
GLCM	Grey level co-occurrence matrix
JA	Jaya algorithm
J-SSO	Jaya-based shark smell optimization
KPA	Known plaintext attack
KP-ABE	Key policy attribute-based encryption
MIQP	Mixed-integer quadratic programming
PRE	Proxy re-encryption
PPX-AC	Privacy-preserving XACML-based access control model
PCA	Principal component analysis
SE	Searchable encryption
SMC	Secure multiparty computation
SSO	Shark smell optimization

ment methodologies, a hybrid cloud concept is introduced which integrates two or more clouds. Private virtual cloud is the term used to describe the common pool of resources in a cloud system [10]–[12].

Security systems, key distribution, encrypting, access, identity authentication, audit scheduling, as well as human and physical access control are some of the security issues affecting cloud data [13], [14], as defined by several privacy-protection strategies [15]–[17]. Improved data security levels have been achieved in the cloud through the development of a privacy-conscious access control system. Researchers have proposed a unique method for combining spectral band handprint images, depending upon the complex dual-tree transform as well as a feature extraction and minimization method based on the Gabor wavelet transform (GWT) and principal component analysis (PCA).

With such an environment taken into consideration, this article offers the following contributions:

– it employs a modified apriori approach to sanitize system information,
– it proposed the CIAO-TME method to generate the best key while adhering to the degree of modification (DOM), information preservation ratio (IPR), and hiding ratio (HR) requirements.

The rest of the paper is as organized follows. Section 2 evaluates the existing literature. The suggested model and its features are described in Section 3. Section 4 covers key generation processes and the data sanitization method. Results and discussion are given while Section 5 while conclusions are presented in Section 6.

2. Literature Review

Danish *et al*. [18] formulated their approach to privacy protection in the cloud environment by employing artificial intelligence (AI). The authors stated that AI abilities were helping companies achieve greater productivity in the corporate cloud environment. Data normalization and recovery were the two key steps of the recommended privacy-preserving system. The extraction of many co-functions, including such factors as DOM, HR, and IPR, leads to the generation of an optimal key. The study demonstrated the effectiveness of the suggested model in improving cloud security when compared to other methods.

Avijit and Radha in [19] employed the honeypot algorithm for data protection or IDS, which was a good strategy for predictions and privacy protection. The dataset was initially standardized using the normalization approach, a process which involves replacing missing values and removing unnecessary data. Following that, unique features were extracted and the best models were chosen using the GLCM algorithm. Predicting the target was done using a unique CNN classifier that offers high attack detection accuracy levels. The developed algorithm was used to protect information from infiltration and other assaults. In addition, a cryptographic mechanism was utilized to ensure the required secrecy protection, while encryption was performed using HPA. If the data holder requests a specific file, the cloud server generates a key and verifies it by interacting with the user for authentication purposes. The performance of the solution was evaluated and was compared with that of other, existing methodologies to demonstrate efficacy of the proposed scheme.

Tehsin *et al*. [20] investigated the privacy-preserving authorization paradigm for the cloud and privacy-preserving strategies for cloud-based EHRs using defined taxonomy. Inner login control and outer security in outsourced system design for hybrid cloud have been formulated and then the PPX-AC algorithm was developed, combining fine login control with the multifunctional use of EHRs and a cutting-edge privacy mechanism. Using HLPN, the authors confirmed the efficiency of the proposed PPX-AC by invalidating known privacy threats. Furthermore, the described model demonstrates its efficacy and multifunctional application possibilities.

Tian *et al*. [21] proposed an IM-based methodology for using MIQP to solve the energy management problem. The viability and efficiency of using IM in MIQP were demonstrated by incorporating cloud-edge architecture. To better fit real applications, the general criteria of the MIQP IM were expanded in terms of security and implementation cost.

Luis *et al*. in [22] presented smart CAMPP data to achieve cloud authorization. Format-conserving encoding methods were used to outsource them discreetly. Furthermore, the observations demonstrated the applicability of the proposed technique, allowing to expect high accuracy levels. In contrast to a method that does not improve security, the authors' proposal has no substantial influence on encryption.

Pan in [23] presented various cloud privacy security vulnerabilities and then proposed a complete privacy security prevention architecture. The characteristics of several techniques were also compared, including network access tools, CP-ABE, KP-ABE, the fine-grain, polynomial number of authority, dismissal mechanism, the detect mechanism, PRE, various leveled encryption, and a mixture of other methods.

Chen *et al*. in [24] designed a lightweight encryption system that maintains an acceptable usefulness model while ensuring evidential privacy preservation. Using the specified prototype system, the recommended approach was deemed secure against a sincere but inquisitive host and a catastrophic collision. The effectiveness of the method was examined and compared to similar solutions using the MNIST and UCI human action recognition database. This strategy decreased the runtime by 20% and the communicated cipher text length by 85%, on average, while maintaining the accuracy of competitive SMC methods.

Yong *et al*. [14] presented a blockchain-based EHR sharing mechanism ensuring both secure and private features by employing encryption algorithms. Additionally, evidence of permission was intended to serve as the consensus protocol for consortium blockchains in order to guarantee the software's reliability. According to a study, the proposed protocol meets the security goals and has a good computational efficiency.

Table 2 summarizes the research on recent cloud data security methods.

Tab. 2. Summary on research in existing papers.

Paper	Proposed metods	Features	Drawbacks
Danish *et al*. [18]	J-SSO algorithm	It is an algorithm with fewer steps and strong global findings	When it comes to ensuring the security of every database, the strategies OI-CSA and BS-WOA strategies offer low convergence performance
Avijit and Radha [19]	Honeypot method	Due to efficient implementation of the system, the security rate of online services has been improved	It is only capable of detecting direct assaults
Tehsin *et al*. [20]	PPX-AC model	Enables fine-grained access control while maintaining privacy	When data is sent between parties, it is essential to make sure that it is secure, as the parties involved were unaware of the information shared between both original parties
Tian *et al*. [21]	MIQP algorithm	Is an optimum solution with good accuracy, while lowering the computational cost	Preserving privacy while decreasing computing costs to the highest possible extend is hard to achieve
Luis *et al*. [22]	Smart CAMPP algorithm	Builds a CA approach for smartphones	No positive feedback on motion sensors leads to challenging results
Sun [23]	CP-ABE, KP-ABE	CP-ABE allows data owners to exchange encryption with authenticated persons through cloud storage while keeping access control settings hidden. Senders can encrypt communications using KP-ABE with a characteristics set	The degree of privacy protection provided by sharing algorithms concerned with users' identities has to be improved
Chen *et al*. [24]	HFWP	It improves training efficiency and communication overhead while providing excellent privacy protection in a short span of time	It cannot gather raw data from consumers
Yong *et al*. [14]	EHR method	Data sharing method among medical institutions were identified	The developer fails to provide timely updates

Distributed locations with cloud processing infrastructures and mass data storage create privacy concerns. For instance, Google's cloud servers are spread out all over the world, including seven sites in the Americas, two in Asia, and three in Europe. Additionally, customers must be aware where the cloud hosting is located, because privacy defining laws vary in many countries. This shows the importance of cyber security aspects in cloud computing.

Many algorithms were introduced to solve the privacy preservation problem, but there is no optimal solution yet. For example the J-SSO algorithm [18] offers a poor convergence rate that causes failures in this sort of applications. Similarly, such algorithms as the honeypot method [19] are capable only of detecting direct assaults, while the smart CAMPP algorithm [23] often creates challenges, as no previous efforts has focused on motion sensors. Moreover, such encryption standards as KP-ABE and CP-ABE also need some improvement in the degree of privacy provided. Therefore, there is still a need of researching the advanced privacy preservation models.

3. Proposed Method Concerning Data Security in Cloud

This work proposes two procedures that attempt to develop a new cloud data security method: data sanitization and restoration process. The data sanitization process sustains the security of sensitive data in the cloud by concealing it from unauthorized users and preventing it from being accessed. In this scenario, data is sanitized using an upgraded version of the a priori algorithm, then data restoration is incorporated for restoring or recovering the sanitized data. In both processes, key generation plays a very important role and should be performed optimally to guarantee data safety. Here, the recommended CIAO-TME technique is used to identify the best key for both data restoration and sanitization sages. Considering this to be an optimization problem, DOM, IPR, and HR-related objectives are used for obtaining the key.

Figure 1 shows the diagram of the concept proposed for ensuring preservation of data security in the cloud.

3.1. Data Sanitization

The association rule mining (ARM) approach is enhanced by using the apriori algorithm containing the following three steps:

1) Scan transaction database once and get the sampling method for each item. During the sampling procedure, the algorithm picks a random sample S from the database da and then searches for frequent item sets S. This can be reduced by lowering the so-called min-support. The

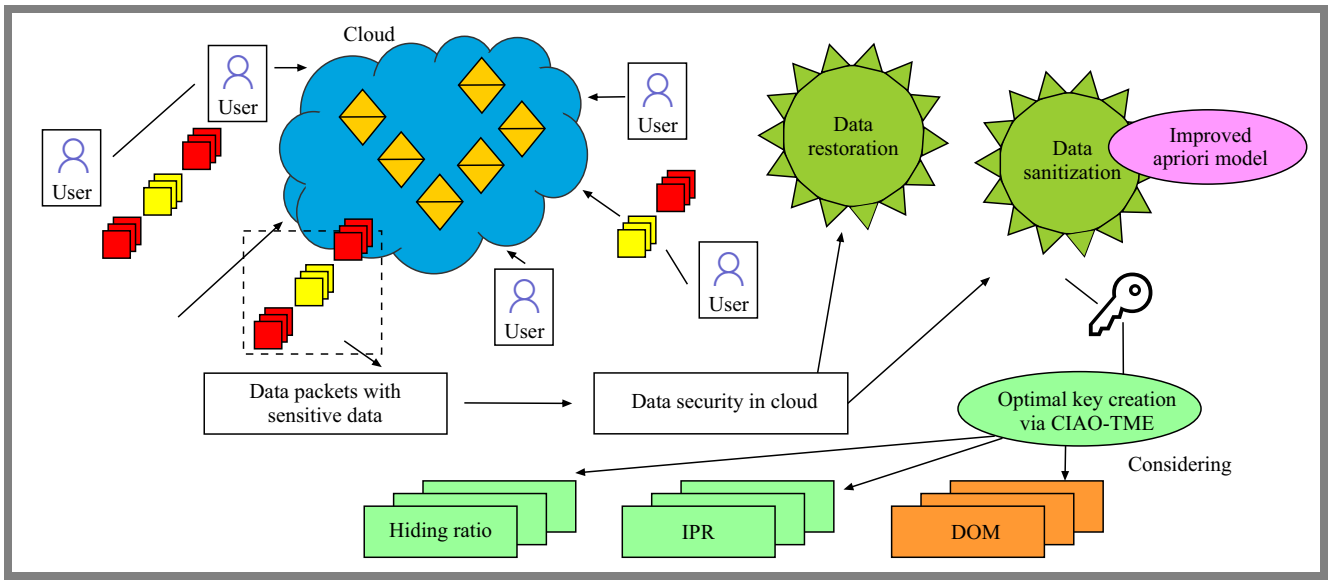

Fig. 1. The proposed concept ensuring data security in the cloud.

impact factor (IF) is found for removing the victim item set. It is fixed with a threshold value $\alpha_{\min}$ such as:

- if $\alpha_{\min}$ or IF is 1 or more, the item set should be continued (considered as sensitive data),
- if $\alpha_{\min}$ or IF equals 0 or less, the item set is considered as victim item set and it should be removed.

2) Using the overlap strategy for counting the support of candidate item set c_k, the sampling sets of L_{k-1} and of L_1 are created.

3) If the $|L_k| \leqslant k$, the algorithm is terminated.

By using the apriori model, the rules for sanitization are created and, similarly, the reverse rules are designed. Next, using the XOR function with the key values, the adopted data sanitization process is performed.

3.2. Key Generation

One of the important and computational time-consuming steps in the creation of a security mechanism is key generation. The creation of uncrackable and non-derivable secure keys is a complicated computational task. Creation of the key matrix and the initial cloud data are essential for preparing sanitized data, such as:

$$d'_S = d_S \oplus Ky_2 \,, \tag{1}$$

where an optimally produced key is referred to as Ky_2, d_S is the original data and d'_S is the sanitized data.

Generating the key is a part of the CIAO-TME model.

3.3. Restoration Procedure

Data sanitization is a technique used for concealing private or sensitive information in a cloud with the goal of preventing unwanted data leakages. Using the special key created during the data cleaning procedure, sensitive data is revealed during data restoration. The opposite is the data restoration process. The same key that the created the CIAO-TME model for the purpose of generating sanitized data is used to recover the original information:

$$\widehat{d}_S = d'_S \oplus Ky_2 \,, \tag{2}$$

where $\widehat{d}_S$ denotes the recovered data.

4. Novel CIAO-TME Optimal Key Selection

To identify the optimal key selection method, first the objective function is formulated as:

$$Obj = \text{Min}\big[\text{DOM} + (1 - \text{HR}) + (1 - \text{IPR})\big] \,, \tag{3}$$

where HR, DOM, and IPR are the objective functions considered for data sanitization. Figure 2 presents the solution encoding scheme.

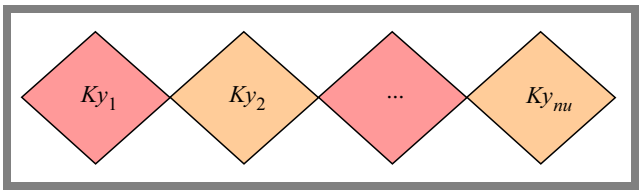

Fig. 2. The proposed data security in the cloud.

HR is the percentage of sensitive items that are properly concealed by d'_S:

$$\text{HR} = \frac{Nd}{tp} \,, \tag{4}$$

where tp is a total number of hidden data indexes, and Nd represents the size of non-zero indexes.

IPR is the inverse of data lost and the rate of non-sensitive rules not concealing in sanitized data set. It is formulated as:

$$\text{IPR} = \frac{1}{tp} N2 \,, \tag{5}$$

where tp is the total number of saved data indexes and $N2$ equals the total amount of zero indexes.

The original dataset d_S and the sanitized dataset d'_S allow to determine the DOM function that is the Euclidean distance between d_S and d'_S:

$$\mathrm{DOM} = -d'_S + d_S \ , \tag{6}$$

where d_S indicates the degree of modification taking place within the unique dataset and d'_S denotes the sanitized dataset d'.

4.1. Proposed CIAO-TME Model

The proposed CIAO-TME model is a hybrid optimization approach and is created by combining the traditional AOA [25] and CHIO techniques [26]. AOA is a high-performance optimization technique as far as convergence time and exploration-exploitation balance are concerned. The advantages of CHIO include high effectiveness when dealing with a large number of optimization issues across a wide range of optimization domains. To overcome the limitations of AOA and CHIO models, the hybrid optimization CIAO-TME model is proposed. It combines two common optimization models to speed up the convergence process offered by solutions [27]–[30]. The individual steps of the model are presented below.

In the first step, a population of N the search agents is initialized, and next the positions of the search agent and other algorithmic parameters are initialized:

$$S^I = LB^I + \mathrm{Rand}(UB^I - Lb^I) \ , \tag{7}$$

where LB^I and UB^I denote the lower and upper bounds of I-th the search agent, respectively.

Then, density Den^I and volume Vol^I of the search agents are set randomly, and acceleration A^I of the I-th the search agent is assigned as:

$$A^I = Rand(UB^I - Lb^I) + LB^I \ . \tag{8}$$

The initial population is then evaluated and the search agent characterized by the best fitness levels is selected as: a_{best}, Vol_{best}, and D_{best}. For the next position, density Den^I_{t+1} and volume Vol^I_{t+1} are determined by:

$$Den^I_{t+1} = Rand(-Den^I_t + Den_{best}) + Den^I_t) \ , \tag{9}$$

$$Vol^I_{t+1} = Vol^I_t + Rand(Vol_{best} + Den_{best}) - Vol^I_t) \ , \tag{10}$$

where Den_{best} and Vol_{best} points out to the best density and best volume, respectively. The Rand function points out to a random value with uniformly distribution. The transfer operator TF is modeled as:

$$TF = \mathrm{e}^{\frac{itr - \max^{itr}}{\max^{itr}}} \ , \tag{11}$$

where itr and $\max^{itr}$ are the current and maximum iterations, respectively.

Step 2 is the exploitation phase. If $TF \leqslant 0.5$, then there occurs no collision between the searching agents. The acceleration factor is modeled as:

$$ACC^I_{t+1} = \frac{Den_{mr} + Vol_{mr} + A_{mr}}{Den^I_{t+1} Vol^I_{t+1}} \ , \tag{12}$$

where Den_{mr}, Vol_{mr}, and A_{mr} denote respectively density, volume, and acceleration of random search agents, respectively.

If $TF > 0.5$, the position of the search agent is updated as:

$$X^I_{t+1} = X^{best}_{t+1} + F \cdot C2\, r\, A^{I-Norm}_{t+1} d\big(T \cdot X^{best} - X^{best}_t\big) \ . \tag{13}$$

In CIAO-TME, the position is updated by combining the concepts of AOA and CHIO as:

$$X^I_{t+1} = \lambda(I) \frac{X^{best}_{t+1} + F \cdot C2\, r\, A^{I-Norm}_{t+1}}{\tau^{\frac{1}{M}}} d\big(T \cdot X^{best} - X^{best}_t\big) \ , \tag{14}$$

where λ refers to the force of infection, τ is the transmission rate, and M refers to the birth rate percentage of individuals who are added to the entire population of M, A^{I-Norm}_{t+1} is the normalized acceleration, $C2 = 6$, $T = C3 \cdot TF$ and $C3 = 0.3$ are the constant values, while F is the flag that shows the direction. In addition, r points out the random value that is generated by means of the tent map to improve the convergence rate.

Step 3 defines the proposed exploration phase. When $TF > 0.5$, then there is no collision and the acceleration of the search agent $itr + 1$ is computed as:

$$ACC^I_{t=1} = \frac{Vol_{best} A_{best} + Den_{best}}{Vol^I_{t+1} Den^I_{t+1}} \ . \tag{15}$$

For CIAO-TME, the acceleration factor is computed using on sample variance S^2 and weight we^i as:

$$ACC^I_{t=1} = \frac{Den_{best} + Vol_{best} A_{best}}{Den^I_{t+1} \frac{Vol^I_{t+1}}{S^2}} we^i \ . \tag{16}$$

For $TF \leqslant 0.5$, the I-th search agent $itr + 1$ is computed as:

$$X^I_{t+1} C1\, r\, d\, A^{I-Norm}_{t+1} \left(X^{rand} - X^I_t\right) + X^{best}_{t+1} \ . \tag{17}$$

As a result, the algorithm in step 4 returns the best solution found.

5. Results and Discussion

The proposed CIAO-TME data security model was implemented in Python using a sample dataset downloaded from [31]. The databases known as Cleveland, Hungary, Switzerland, and VA Long Beach were used. The accessible model was evaluated over AOA, CHIO, HBA, BES, and BMO for a variety of metrics, such as DOM, HR, IPR, and cost.

5.1. Convergence Analysis

Figure 3 shows the cost factors for three datasets (D1, D2, and D3), illustrating convergence of the implemented CIAO-TME and comparing it with the traditional schemes: AOA, CHIO, HBA, BES, and BMO. CHIO and AOA have revealed poor performance by acquiring more expensive assets in the first, second, and third scenarios. From the 12-th to the 50-th

iteration, CIAO-TME achieves a minimum convergence of 0.001 for all scenarios and slightly higher cost values before the 10-th iteration.

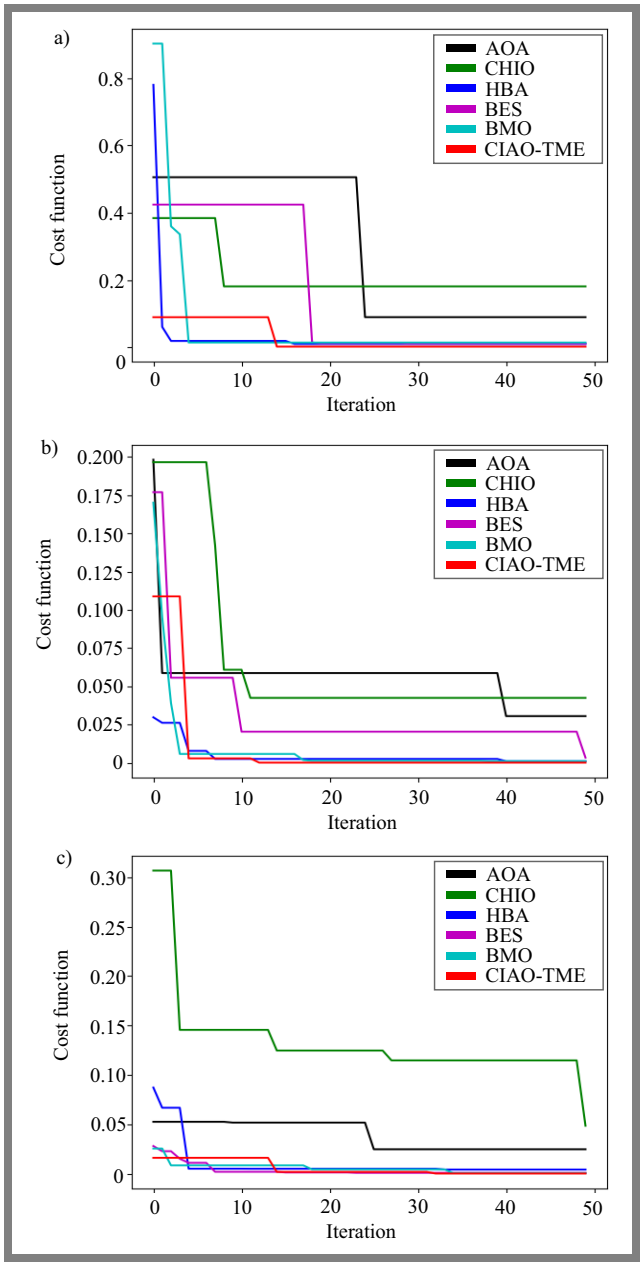

Fig. 3. Convergence of CIAO-TME for: a) D1, b) D2, and c) D3 datasets.

5.2. Analysis of DOM, HR, and IPR

Figure 4 illustrates the analysis of DOM, HR, and IPR parameters for the CIAO-TME method over known techniques, for D1, D2, and D3 datasets. The proposed CIAO-TME scheme achieved the lowest DOM (∼0.45) – a result that better than the one characterizing other schemes, such as AOA, CHIO, HBA, BES, and BMO. Additionally, CIAO-TME achieved a higher HR (∼1.0) for D2 and D3 and the lowest HR for D1, when compared to D2 and D3. Moreover, the proposed CIAO-TME scheme obtained a better IPR

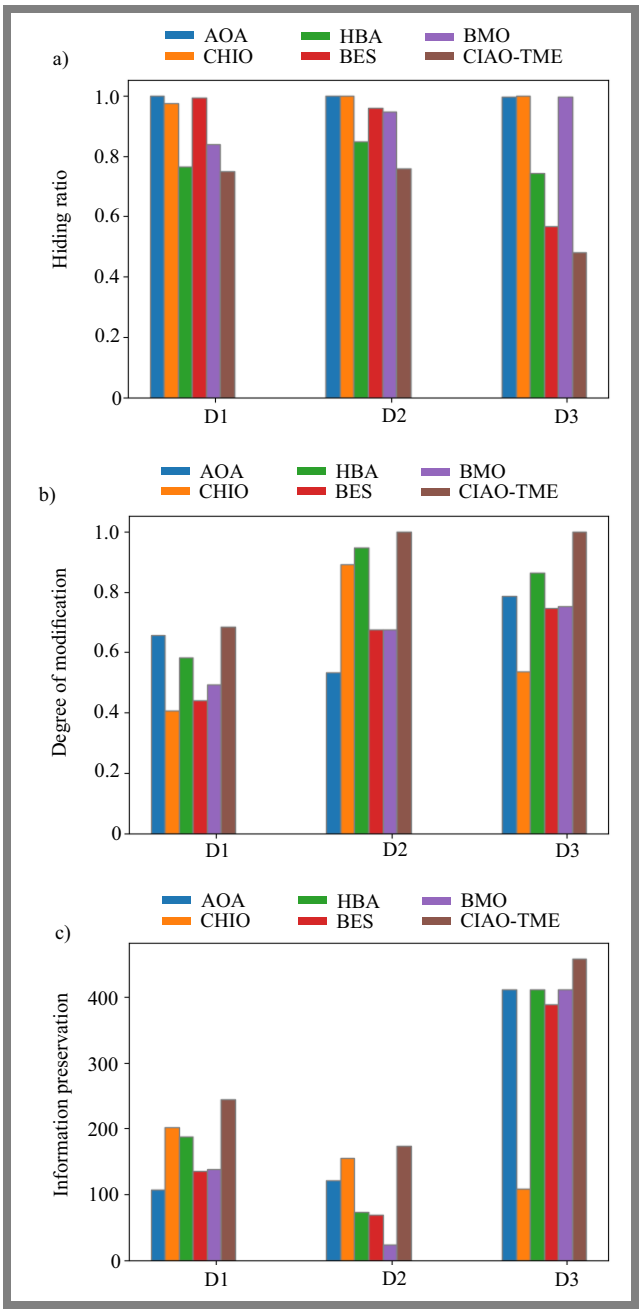

Fig. 4. Analysis of known schemes versus CIAO-TME for: a) DOM, b) HR, and c) IPR.

(450) for D3 and a relatively low IPR for D1 and D2, when compared to D3. The outcomes of using CIAO-TME are better when compared with AOA, CHIO, HBA, BES, and BMO models, especially with regard to the D3 dataset, due to the optimal generation of the key and an improved apriori-based ARM.

5.3. Fitness Analysis

Figure 5 shows the fitness examination defined by Eq. (3) for the CIAO-TME scheme over AOA, CHIO, HBA, BES, and BMO for D1, D2, and D3 datasets. The objective is to minimize the HR function while keeping IPR high and DOM at a relatively low levels. Here, the minimal outcomes

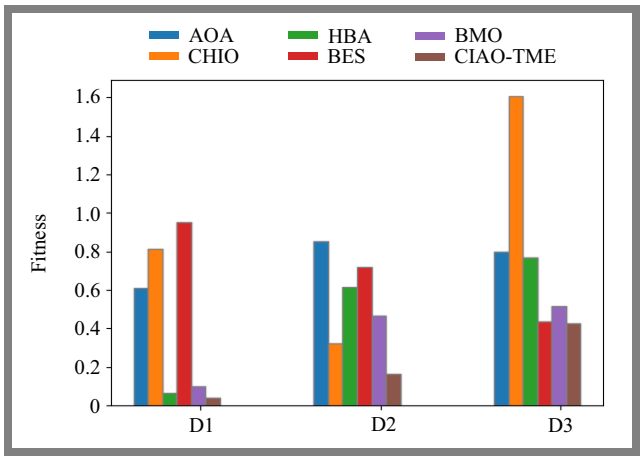

Fig. 5. Fitness analysis of CIAO-TME compared with others methods for D1, D2, and D3 datasets.

are achieved by CIAO-TME at D3. For D1 and D2, the CIAO-TME model has attained the best outcomes in terms of the objective. However, compared to D1, D2, D3-related outcomes were the best. This is the result of optimal generation of the key and an improved apriori-based ARM.

5.4. Comparison Summary

The proposed CIAO-TME cloud data security scheme is compared with existing ARM schemes for different metrics (Table 3). CIAO-TME-based security models are tested in terms of performance fitness at high HR, high IPR, and low DOM. An HR of 0.68 is achieved for CIAO- TME, i.e. a result that is better than the one obtained when using ARM or an approach with no CIAO-TME involved. The best results were achieved on D1 and D2 datasets. The merging of optimal key generation and the improved apriori-based ARM additionally enhanced the CIAO-TME scheme.

Tab. 3. Comparison of the proposed scheme with other rule-mining methods and metrics.

D1 dataset			
Metrics	**CIAO-TME**	**Proposed with existing ARM**	**Proposed with no CIAO-TME**
Fitness	0.0429	0.8012	0.1987
HR	0.6859	0.5573	0.5573
IPR	244	108	108
DOM	0.7486	1	1
D2 dataset			
Metrics	**CIAO-TME**	**Proposed with existing ARM**	**Proposed with no CIAO-TME**
Fitness	0.1638	0.597	1.097
HR	1	0.9324	0.7891
IPR	174	5	54
DOM	0.7589	1	1
D3 dataset			
Metrics	**CIAO-TME**	**Proposed with existing ARM**	**Proposed with no CIAO-TME**
Fitness	0.4263	0.9637	0.5495
HR	1	0.8624	0
IPR	458	0	411
DOM	0.4823	1	0.9951

5.5. Analysis of Existing Works

For three test instances, Table 4 shows the comparison of the selected CIAO-TME with other existing schemes, such as GMGW [31] and J-SSO [18]. The adopted CIAO-TME-based security method is evaluated in terms of such metrics as fitness, HR, IPR, and DOM, with the lowest DOM (~ 0.48) achieved for D3. This result is superior to that of other currently used techniques. The D2 and D3 datasets allowed to obtained an improved HR ($\sim$1.0) by using CIAO-TME, and for the D1 dataset, the proposed solution achieved the lowest HR ($\sim$0.68). Additionally, the IPR on D3 was higher ($\sim$458), while for D1 and D2 IPR it was relatively low. Overall, the CIAO-TME produces superior results than AOA, CHIO, HBA, BES, and BMO models.

Tab. 4. Comparison of CIAO-TME with other methods.

D1 dataset			
Metrics	**CIAO-TME**	**GMGW [31]**	**J-SSO [18]**
Fitness	0.0429	0.7554	0.1023
HR	0.6859	0.3286	0.513
IPR	244	202	205
DOM	0.7486	0.975	0.9639
D2 dataset			
Fitness	0.1638	1.663	0.3699
HR	1	0.8531	0.647852
IPR	174	148	74
DOM	0.7589	1	0.8486
D3 dataset			
Fitness	0.4263	1.024	0.7383
HR	1	0.8122	0.8864
IPR	458	388	411
DOM	0.4823	1	0.7435

5.6. Attack Analysis

Tables 5–6 present the examination results for three datasets under simulated CPA and KPA attacks. A CPA is a cryptanalysis attack paradigm that assumes the attacker has access to the cypher texts for any plain. The KPA is a cryptanalysis attack type in which the attacker has access to both plaintext (also known as a crib) and encrypted version of the data (cipher text). These can be used to divulge additional classified information, including security codes and private keys. CIAO-TME attained minimal KPA attack and CPA attack vulnerability values versus AOA, CHIO, HBA, BES, and BMO for all datasets. Specifically, the third datasets revealed lower attack parameter values for the CPA, while for the KPA, dataset D1 is characterized by lower attack values than datasets D2 and D3. This significant improvement stems from due the enhanced apriori-based ARM and optimized key generation.

Tab. 5. CPA analysis results.

Methods	AOA	CHIO	HBA	BES	BMO	CIAO-TME
Dataset D1	0.1975	0.2834	0.2346	0.2701	0.2212	0.1686
Dataset D2	0.2125	0.2434	0.2653	0.2275	0.4494	0.1769
Dataset D3	0.2173	0.3794	0.3917	0.2928	0.2715	0.1629

Tab. 6. KPA analysis of the proposed and other known schemes.

Methods	AOA	CHIO	HBA	BES	BMO	CIAO-TME
Dataset D1	0.1862	0.3429	0.3003	0.3275	0.2793	0.1197
Dataset D2	0.1658	0.2658	0.2527	0.2049	0.2274	0.1285
Dataset D3	0.1835	0.2572	0.2713	0.386	0.2237	0.1612

6. Conclusion

This work proposes two procedures to protect sensitive data from unauthorized access and users by relying on the data sanitization process. The modified apriori approach is used to clean the data and the major goal is to generate the best key – a task accomplished by employing an optimization method. A reversible method, known as data restoration, allows to retrieve or obtaining cleaned content. IPR, HR, and DOM were the objectives set to tackle the optimization problem. In the simulations, the CIAO-TME scheme achieved the lowest DOM for the D3 dataset and overperformed other techniques that are currently in use, such as J-SSO and GMGW. The D2 and D3 datasets improved HR to approx. 1.0 with the CIAO-TME scheme. The CIAO-TME also achieved the lowest HR (0.68) on the D1 dataset.

References

[1] Y. Wang, A. Zhang, P. Zhang, and H. Wang, "Cloud-assisted EHR sharing with security and privacy preservation via consortium blockchain", *IEEE Access*, vol. 7, pp. 136704–136719, 2019 (https://doi.org/10.1109/ACCESS.2019.2943153).

[2] X. Yang, M. Wang, X. Wang, G. Chen, and C. Wang, "Stateless cloud auditing scheme for non-manager dynamic group data with privacy preservation", *IEEE Access*, vol. 8, pp. 212888–212903, 2020 (https://doi.org/10.1109/ACCESS.2020.3039981).

[3] H. Liu, X. Yao, T. Yang, and H. Ning, "Cooperative privacy preservation for wearable devices in hybrid computing-based smart health", *IEEE Internet of Things Journal*, vol. 6, no. 2, pp. 1352–1362, 2019 (https://doi.org/10.1109/JIOT.2018.2843561).

[4] H. Yan and W. Gui, "Efficient identity-based public integrity auditing of shared data in cloud storage with user privacy preserving", *IEEE Access*, vol. 9, pp. 45822–45831, 2021 (https://doi.org/10.1109/ACCESS.2021.3066497).

[5] C. Xu, N. Wang, L. Zhu, K. Sharif, and C. Zhang, "Achieving searchable and privacy-preserving data sharing for cloud-assisted e-healthcare system", *IEEE Internet of Things Journal*, vol. 6, no. 5, pp. 8345–8356, 2019 (https://doi.org/10.1109/JIOT.2019.2917186).

[6] B.A. Jalil, T.M. Hasan, G.S. Mahmood, and H.N. Abed, "A secure and efficient public auditing system of cloud storage based on BLS signature and automatic blocker protocol", *Journal of King Saud University – Computer and Information Sciences*, vol. 34, no. 7, pp. 4008–4021, 2022 (https://doi.org/10.1016/j.jksuci.2021.04.001).

[7] K. Wang, C.-M. Chen, Z. Tie, M. Shojafar, S. Kumar, and S. Kumari, "Forward privacy preservation in IoT-enabled healthcare systems", *IEEE Transactions on Industrial Informatics*, vol. 18, no. 3, pp. 1991–1999, 2022 (https://doi.org/10.1109/TII.2021.3064691).

[8] Q. Kong, R. Lu, F. Yin, and S. Cui, "Privacy-preserving continuous data collection for predictive maintenance in vehicular fog-cloud", *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 8, pp. 5060–5070, 2021 (https://doi.org/10.1109/TITS.2020.3011931).

[9] J. Wang, D. Shi, J. Chen, and C.-C. Liu, "Privacy-preserving hierarchical state estimation in untrustworthy cloud environments", *IEEE Transactions on Smart Grid*, vol. 12, no. 2, pp. 1541–1551, 2021 (https://doi.org/10.1109/TSG.2020.3023891).

[10] M. Fernandes, J. Decouchant, M. Völp, F.M. Couto, and P. Esteves-Verissimo, "DNA-SeAl: Sensitivity levels to optimize the performance of privacy-preserving DNA alignment", *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 3, pp. 907–915, 2020 (https://doi.org/10.1109/JBHI.2019.2914952).

[11] R. Li, T. Song, B. Mei, C. Hu, W. Li, M. Larson, X. Cheng, and R. Bie, "A cloud-based framework for verifiable privacy-preserving spectrum auction", *High-Confidence Computing*, vol. 2, 2022 (https://doi.org/10.1016/j.hcc.2021.100037).

[12] S. Mewada, "Data mining-based privacy preservation technique for medical dataset over horizontal partitioned", *International Journal of E-Health and Medical Communications* (IJEHMC), vol. 12, no. 5, pp. 50–66, 2021 (https://doi.org/10.4018/IJEHMC.20210901.oa4).

[13] X. Xu, S. Fu, L. Qi, X. Zhang, Q. Liu, Q. He, and S. Li, "An IoT-oriented data placement method with privacy preservation in cloud environment", *Journal of Network and Computer Applications*, vol. 124, pp. 148–157, 2018 (https://doi.org/10.1016/j.jnca.2018.09.006).

[14] C. Fang, Y. Guo, N. Wang, and A. Ju, "Highly efficient federated learning with strong privacy preservation in cloud computing", *Computers & Security*, vol. 96, Article 101889, 2020 (https://doi.org/10.1016/j.cose.2020.101889).

[15] K.M. Prabha and P.V. Saraswathi, "Suppressed K-anonymity multi-factor authentication based Schmidt-Samoa cryptography for privacy preserved data access in cloud computing", *Computer Communications*, vol. 158, pp. 85–94, 2020 (https://doi.org/10.1016/j.comcom.2020.04.057).

[16] C.-T. Li, D.-H. Shih, and C.-C. Wang, "Cloud-assisted mutual authentication and privacy preservation protocol for telecare medical information systems", *Computer Methods and Programs in Biomedicine*, vol. 157, pp. 191–203, 2018 (https://doi.org/10.1016/j.cmpb.2018.02.002).

[17] J. Mandala and M.V.P. Chandra Sekhara Rao, "Privacy preservation of data using crow search with adaptive awareness probability", *Journal of Information Security and Applications*, vol. 44, pp. 157–169, 2019 (https://doi.org/10.1016/j.jisa.2018.12.005).

[18] D. Ahamad, S.A. Hameed, and M. Akhtar, "A multi-objective privacy preservation model for cloud security using hybrid Jaya-based shark smell optimization", *Journal of King Saud University – Computer and Information Sciences*, vol. 34, no. 6, part A, pp. 2343–2358, 2020 (https://doi.org/10.1016/j.jksuci.2020.10.015).

[19] A. Mondal and R.T. Goswami, "Enhanced HoneyPot cryptographic scheme and privacy preservation for an effective prediction in cloud security", *Microprocessors and Microsystems*, vol. 81, 2021 (https://doi.org/10.1016/j.micpro.2020.103719).

[20] X. Yao, F. Farha, R. Li, I. Psychoula, L. Chen, and H. Ning, "Security and privacy issues of physical objects in the IoT: Challenges and opportunities", *Digital Communications and Networks*, vol. 7, no. 3, pp. 373–384, 2021 (https://doi.org/10.1016/j.dcan.2020.09.001).

[21] T. Kanwal, A. Anjum, S.U.R. Malik, A. Khan, and M.A. Khan, "Privacy preservation of electronic health records with adversarial attacks identification in hybrid cloud", *Computer Standards & Interfaces*, vol. 78, 2021 (https://doi.org/10.1016/j.csi.2021.103522).

[22] N. Tian, Q. Guo, and H. Sun, "Privacy preservation method for MIQP-based energy management problem: A cloud-edge framework", *Electric Power Systems Research*, vol. 190, 2021 (https://doi.org/10.1016/j.epsr.2020.106850).

[23] L. Hernández-Álvarez, J. María de Fuentes, L. González-Manzano, and L.H. Encinas, "SmartCAMPP – Smartphone-based continuous authentication leveraging motion sensors with privacy preservation", *Pattern Recognition Letters*, vol. 147, pp. 189–196, 2021 (https://doi.org/10.1016/j.patrec.2021.04.013).

[24] P.J. Sun, "Security and privacy protection in cloud computing: Discussions and challenges", *Journal of Network and Computer Applications*, vol. 160, 2020 (https://doi.org/10.1016/j.jnca.2020.102642).

[25] F. A. Hashim, K. Hussain, E.H. Houssein, M.S. Mabrouk, and W. Al-Atabany, "Archimedes optimization algorithm: a new metaheuristic algorithm for solving optimization problems", *Applied Intelligence*, vol. 51, pp. 1531–1551, 2021 (https://doi.org/10.1007/s10489-020-01893-z).

[26] F. Martínez-Álvarez, G. Asencio-Cortés, J.F. Torres, D. Gutiérrez-Avilés, L. Melgar-García, R. Pérez-Chacón, C. Rubio-Escudero, J.C. Riquelme, and A. Troncoso, "Coronavirus optimization algorithm: A bioinspired metaheuristic based on the COVID-19 propagation model", *Big Data*, vol. 8, no. 4, 2020 (https://doi.org/10.1089/big.2020.0051).

[27] M.M. Beno, I.R. Valarmathi, S.M. Swamy, and B.R. Rajakumar, "Threshold prediction for segmenting tumour from brain MRI scans", *International Journal of Imaging Systems and Technology*, vol. 24, no. 2, pp. 129–137, 2014 (https://doi.org/10.1002/ima.22087).

[28] R. Thomas and M.J.S. Rangachar, "Hybrid optimization based DBN for face recognition using low-resolution images", *Multimedia Research*, vol. 1, no. 1, pp. 33–43, 2018 (DOI: 10.46253/j.mr.v1i1.a5).

[29] J. Devagnanam and N.M. Elango, "Optimal resource allocation of cluster using hybrid grey wolf and cuckoo search algorithm in cloud computing", *Journal of Networking and Communication Systems*, vol. 3, no. 1, pp. 31–40, 2020 (https://doi.org/10.46253/jnacs.v3i1.a4).

[30] R.A. Mustafa, H.S. Chyad, and J.R. Mutar, "Enhancement in privacy preservation in cloud computing using apriori algorithm", *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 26, no. 3, pp. 1747–1757, 2022 (https://doi.org/10.11591/ijeecs.v26.i3.pp1747-1757).

[31] M.M. Annie Alphonsa and P. Amudhavalli, "Genetically modified glowworm swarm optimization based privacy preservation in cloud computing for healthcare sector", *Evolutionary Intelligence*, vol. 11, pp. 101–116, 2018 (https://doi.org/10.1007/s12065-018-0162-4).

Rajkumar Patil
Assistant Professor, Department of Computer Science Engineering
E-mail: rajkumarpatil8827@gmail.com
Department of Computer Science Engineering, GITAM (Deemed to be) University, Hyderabad, Telangana, India

Gottumukkala HimaBindu, PhD
Assistant Professor, Department of Computer Science Engineering
E-mail: hgottumu@gitam.edu
Department of Computer Science Engineering, GITAM (Deemed to be) University, Hyderabad, Telangana, India