International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Discrete Particle Swarm Optimization based Data Encryption and Distribution to Mass Cloud Storage system

Sasidhar Attuluri[1*], Mona Ramesh[2]

Submitted: 12/09/2023 **Revised:** 28/10/2023 **Accepted:** 12/11/2023

Abstract: During the last several decades, the amount of data created and the number of digital records that exist has grown exponentially due to the widespread use of data-driven applications and unified systems. The need for fast and effective data storage and retrieval is becoming more important as the data volumes handled applications continue to expand exponentially. One of the most effective ways to deal with ever-increasing data loads is to make advantage of cloud storage. Yet, maintaining data privacy while storing information outside on the cloud is a significant difficulty when using this strategy. When it comes to large data, where safeguarding data in a fast and correct way is a hard effort, data confidentiality is one of the main issues connected with cloud storage, which has contributed heavily as an impediment for cloud computing adoption throughout the globe. Our research intends to aid in the fight against cybercrime by ensuring the privacy of crucial agricultural data. We provide an architecture for protecting privacy while sharing data over the cloud, one that can scale with your data in both time and space. Combining the Particle Swarm Optimization (PSO) algorithm with parallel data dissemination and selective encryption to protect user privacy is fundamental to our architecture. Experiments and analysis of competing frameworks show that our suggested one is superior in terms of execution time and memory footprint.

Keywords: Data encryption, parallel distribution of data, Discrete Particle Swarm Optimization, cloud storage.

1. Introduction

The amount of digital data and records has been growing at a very fast rate after Industry 3.0. This is due to the widespread adoption and use of IT-applications and services, such as social networking sites, sensor networks, internet healthcare applications, and many others [6]. This has turned databases into "data warehouses," which are used to store and organise "big data." Cloud Computing, Business Intelligence, and Data Mining are some of the new technologies that are being used a lot to handle big data [7]. But the problem of how to store and get data quickly is still a big one [8, 9]. This is especially true for applications that use a lot of data. To deal with this problem, companies and research-based organisations are now switching to Cloud Computing, a relatively new type of computing environment that offers fast results and a number of other services. Cloud computing gives its users a number of useful services based on their needs, which affects the cost and performance of the service as a whole. One of its most popular services is called Cloud Storage Service or just Cloud Storage. This service lets users access their shared data and internet resources from anywhere and at any time, which is very convenient [10].

By definition, it is a data storage unit owned and operated by a Cloud Service Provider (CSP) [11] in which digital data is kept in logical pools across numerous computers (and frequently locations). As cloud storage enables customers to rent virtual storage proportionate with their needs and pay correspondingly for the services, it is being utilised widely to serve the goal of storing the enormous mega-form of data known as Big-Data [12]. With the assurance that their data will always be accessible when they need it, clients/companies may move away from conventional computer environments and into the cloud [13]. When a customer uses cloud storage, he is essentially outsourcing his data, which presents a number of hazards and difficulties. The insecurity of our data is one such issue [14]. When it comes to large amounts of sensitive data, cloud storage data security is the greatest and most visible obstacle [15]. Confidentiality, integrity, and availability are the three primary pillars of data security [16].

In this paper, the proposed model is intended to secure the data through effective evolutionary algorithm based encryption mechanism and distribute the data to multiple cloud environment includes Amazon and Microsoft cloud

[1]*Data Protection Officer*
Savin Technologies Inc. |Suite # 2070, 9901 East Valley Ranch Parkway,| Irving, TX 75063
sattuluri@savintechnologies.com
[2]*Research Scholar*
School of Computer Science and Technology
Karunya Institute of Technology and Sciences
Karunya Nagar, Coimbatore, India - 641114
monar@karunya.edu.in

storage systems. The rest of the paper is organized as follows: Part 2 covers prior work in the field of cloud computing data encryption and dissemination. Data distribution modelling and the problem description may be found in Section 3, along with the problem's formulation. The proposed BPSO and the method for creating keys are described in Section 4. In Part 5, we outlined how the experiments would be conducted and analysed; in Section 6, we concluded the research and offered suggestions for future enhancements to the procedure.

2. Related Works

This section describes the many methods that have been developed and deployed for protecting sensitive information. Data encryption is one method proposed by experts to ensure privacy of sensitive information. Encryption of data is a well-known method of addressing privacy issues. Traditional encryption methods may not work well in the cloud because of the need for faster and more memory/space-efficient procedures when dealing with large amounts of data. This critical issue of data confidentiality and security maintenance has prompted a lot of study and a number of proposed solutions.

An activity monitoring system based on data and user signatures generated with the SHA-256 cryptographic hash technique was suggested by Velmurugadass, P. et al. [17] in 2021. The plan encrypts information (in packets) using the Elliptic Curve Integrated Encryption Scheme (ECIES) method before sending it to a remote server in the cloud. In 2021, Thabit et al. [18] developed a hybrid structure symmetric encryption technique using the Festial approach and substitution permutation. It supports data blocks of either 128 or 256 bits in size, and the corresponding symmetric keys are 128 or 256 bits in length. It uses XOR and XNOR, and after 4 passes, satisfies Shannon's theory of diffusion and confusion. In 2020, Samriya, K. J., and Kumar, N. [19] proposed a hybrid strategy to secure cloud-based systems and applications against intrusion, data leakage, phishing assaults, and fake entity detection. The method employs a combination of fuzzy logic and ANN for anomaly grouping across a broad dynamic range.

An end-to-end data security architecture based on Ethereum Blockchain was presented by Ullah, A. et al. [20], (2020) to secure information passing across the various SACADA system levels. Ngnie Sighom et al. [21] (2017) suggested a method for protecting cloud-stored data that combines Advanced Encryption Standard (AES) with Information Dispersal Algorithms and Secure Hash Algorithm-512. Using identity-based authentication (hashed IDs) of participating IoT devices and AES as underlying encryption approach, Zhang et al. [22], (2019) proposed a semi-outsourcing privacy-preserving scheme. Time and communication expenditures are well managed by the plan. To ensure the privacy of sensitive information while allowing for easy access to it, Xu et al. [23], 2018 developed a Ciphertext-Policy Attribute-Based Encryption (CP-ABE) using bilinear pairing. Asymmetric encryption for dynamic encryption with two-factor influenced encoding was suggested by Liu et al. [24], 2015. To protect information, a public key is used. Yet, there were two prerequisites for decryption. a) the owner's private/secret key, or b) the security system's unique passcode.

Mathematical benchmark functions [25], the Artificial Bee Colony algorithm for solving numerical problems [26], node localization in wireless sensor networks [27], multi-objective problems [28], effective node placement in wireless sensor networks [29], data mining [30], and so on are just some of the many areas where evolutionary algorithms have found success. In this study, the inspiration for producing an optimum key for data encryption comes from the optimization tactics that were followed in these streams.

3. Problem Definition

The proposed model is intended to encrypt the data and store it to the secure cloud system in a distributed manner. In this section the basic system architecture that are used for secure encryption and data distribution will be discussed along with the data distribution strategy.

3.1. System Architecture

As Mass Distributed Storage is one of the key strategies utilised by cloud storage to store huge data, our system is likely to outperform conventional encryption methods when applied to cloud storage. It is recommended that the proposed model be implemented on the client-side system to act as a pre-processing phase on cloud storage. The suggested system relies on logically separated functional modules for analysis and processing of gathered data.

3.2. Privacy preservation model on cloud centre

Initially the user data will be split into Non-Sensitive and Sensitive data based on Logistic Regression [1]. And the sensitive data will be further classified into two equal parts namely S_1 and S_2. This data will be encrypted using an encryption key generated by Discrete PSO which will be discussed in Section 4.

The representation of sensitive data split up into two equal parts can be represented as

$$S_1 = \frac{D}{2} \tag{1}$$

where D refers the sensitive data size and S_2 can be computed as

$$S_2 = D \cap S_1 \tag{2}$$

These two halves of the data will be encrypted using Discrete PSO generated optimized key. And the resultant

factor can be further encrypted Trans Key which was generated mathematically as

$$Key_{Trans} = \frac{d(K_{DPSO})}{|D|} \quad (3)$$

where $d()$ represents the total number of digits in the key generated by DPSO.

After generation of Key_{Trans} it will be used to encrypt the data further using XOR operation as follows

$$X'' = Key_{Trans} \oplus X' \quad (4)$$

where $X \in \{S_1', S_2'\}$.

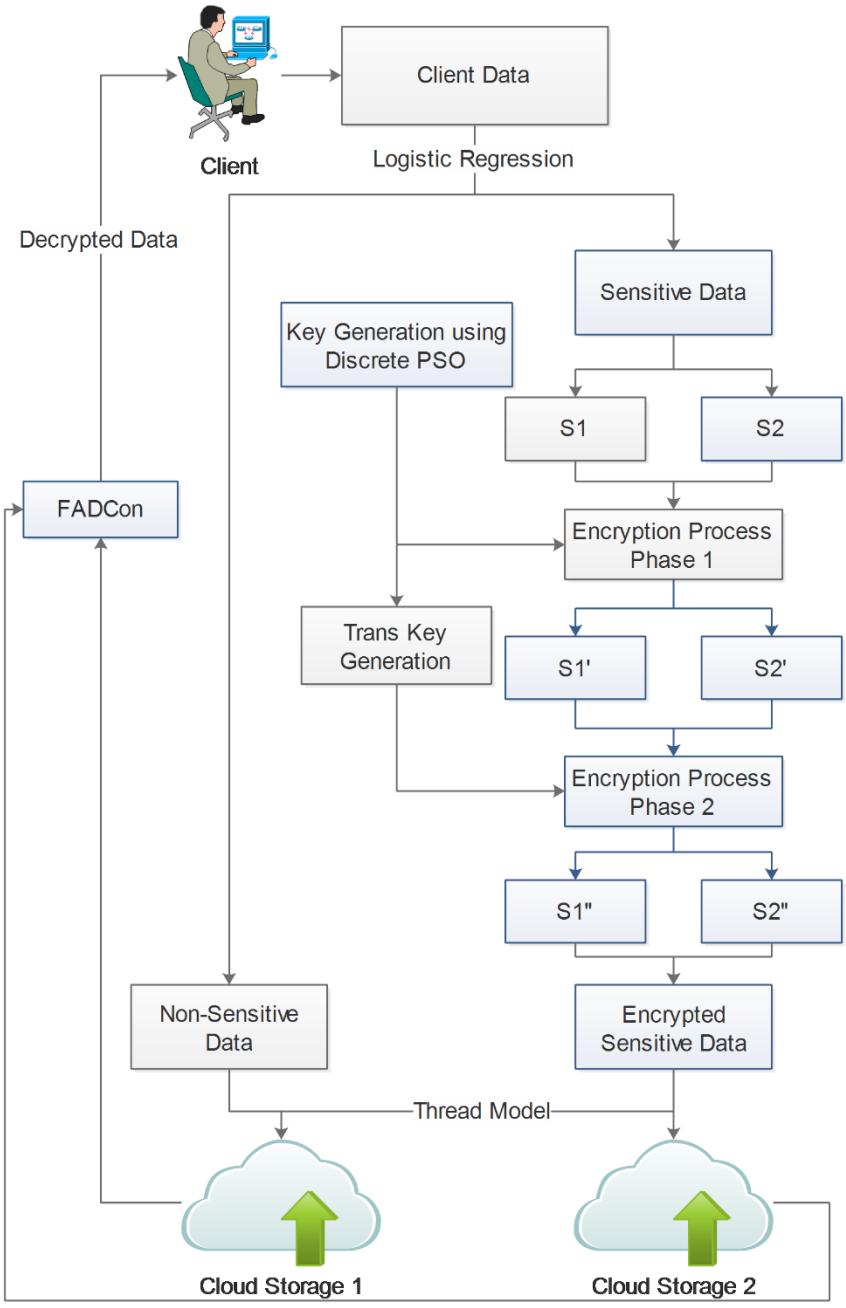

Fig 1. System architecture

4. Proposed System

4.1. Discrete PSO

The key generation is the main procedure that holds the primal part of encryption model. In this proposed model the Particle Swarm Optimization algorithm is used to generate the key based on hiding ratio. The PSO will be formulated as Discrete PSO using Roundoff Mechanism. The pseudocode of the proposed model is given in Algorithm 1.

Algortihm1: Discrete PSO for key generation

Input: Objective Function (Hiding Ration f()), Termination Criteria, Population Size
// Population Initialization
1: for $i = 1\ to\ N$ do
2: $X_i = randperm(1, D_{att})$
3: $F_X^i = f(X_i)$
4: end for
//Iteration Starts
5: while (Termination criteria) do
6: $V_i = w * (C_1 * r * (X_{best} - X_i)) + (C_2 * r * (X_{Gbest} - X_i))$
7: $X_{i+1} = X_i + V_i$
8: $X_{i+1} = Roundoff(X_{i+1})$
9: $F_X^i = f(X_i)$
10: Until (Termination Criteria Satisfied)

Output: X_{Gbest}

4.2. FADCon

For decryotion procedure the output of the mode will be reversed operation of encryption where the encrypted data will be undergoing XOR operation with Trans Key to find the original data.

$$X' = X'' \oplus Key_{Trans} \quad (5)$$

4.3. Data Distribution using Thread Model

For cloud service deployment architectures, cloud servers are often seen as a trusted third-party service supplied by a trusted CSP. Yet, when looking at CSPs that are either partially trusted or not trusted at all, it has been shown that cloud operators are more likely to commit unanticipated confidentiality breaches than malicious assaults [2].

We may conclude that data is not completely safe in terms of secrecy, despite the fact that the activities are limited by the legislation, since cloud providers require access to the data for governance reasons. We establish two threat models based on existing cloud practises, using cloud operator activities as the source of risks **[3].** The two models are namely Anti-Regulatory Compliance and Malicious Access threat model.

5. Experimental Analysis

5.1. Experimental Setup

The performance of the proposed model is evaluated based on the implementation of the proposed model with existing algorithms in Cloud Amazon AWS S3 storage and Microsoft storage units. Two parallel distributed storage units are the above-mentioned storage units. The proposed algorithm has been compared with the existing algorithms such as AES algorithm, SED2[4], EDCon [4], ICA [4] and IDED [5]. The used dataset descriptions are available in Table I.

Table 1. Dataset chosen for evaluation

Sl.No.	Dataset Name	Size	Dataset Type	No. of files
1	**DS1**	1 KB	Small	10
2	**DS2**	1 MB		
3	**DS3**	10 MB		
4	**DS4**	50 MB	Large	
5	**DS5**	250 MB		
6	**DS6**	500 MB		

5.2. Experimental Evaluation

The results of the proposed and existing models are compared on three aspects namely time taken for encryption, decryption and the total memory utilized in terms of Bytes. The interpretation of results are given category wise in this section.

5.2.1 Experimental Evaluation w.r.t encryption time

Table 2: Experimental Results of proposed Vs existing algorithms w.r.t. encryption time

DataSet	AES	SED2	IDED	BPSO
DS1	3.85	4.07	3.45	3.08
DS2	14.22	12.31	10.72	9.72
DS3	40.86	38.20	37.33	33.33
DS4	7736.07	6924.35	6229.87	6208.87
DS5	29935.11	26955.01	23209.91	22292.91
DS6	53686.56	47273.12	40654.54	38443.54

Table 2 shows the encryption time in seconds taken by the proposed and existing algorithm on large and small datasets. Figure 2a and 2b shows the graphical representation of results in Table 2 for better interpretation.

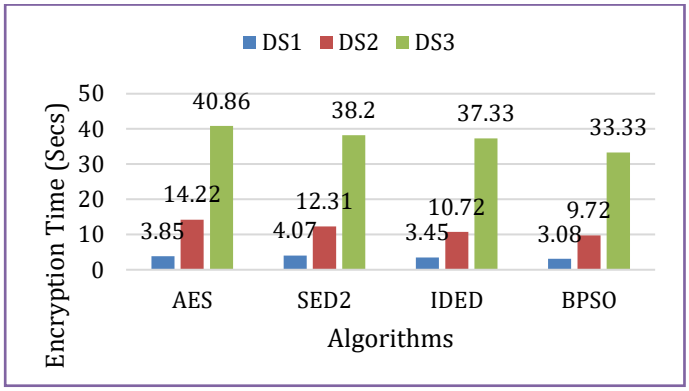

Fig 2a. comparison of results w.r.t. encryption time on small size datasets

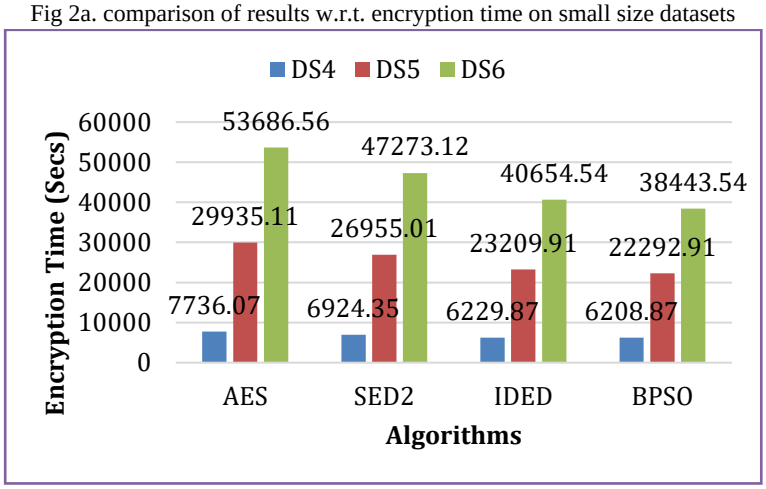

Fig 2b. Comparison of results w.r.t. encryption time on large size datasets

On comparing the results of 2a, on small sized datasets the proposed model outperforms significantly when compared with existing methods w.r.t. encryption time. In specific, on DS1, BPSO outperforms AES with 20%, SED2 with 24.3% and IDED with 10.7%. On DS2, BPSO outperforms AES with 31.6%, SED2 with 21% and IDED with 9.3%. On DS3, BPSO outperforms AES with 18.4%, SED2 with 12.7% and IDED with 10.7%. On comparing the results in Figure 2b, in large datasets DS4 BPSO outperforms AES with 19.7%, SED2 with 10.3% and IDED with 0.3%. On DS5, BPSO outperforms AES with 25.5%, SED2 with 17.3% and IDED with 4%. On DS6, BPSO outperforms AES with 28.4%, SED2 with 18.7% and IDED with 5.7%.

5.2.2 Experimental Evaluation w.r.t encryption time

Table 3. Experimental Results of proposed Vs existing algorithms w.r.t. decryption time

DataSet	AES	EDCon	IDED	BPSO
DS1	3.30	2.52	2.74	2.34
DS2	12.91	12.55	10.41	8.41
DS3	38.28	37.02	34.41	30.41
DS4	6970.13	6706.21	5583.34	5563.34
DS5	28035.22	25141.89	22429.76	21071.76
DS6	54007.11	46840.43	38289.44	35537.44

Table 3 shows the decryption time in seconds taken by the proposed and existing algorithm on large and small datasets. Figure 3a and 3b shows the graphical representation of results in Table 3 for better interpretation.

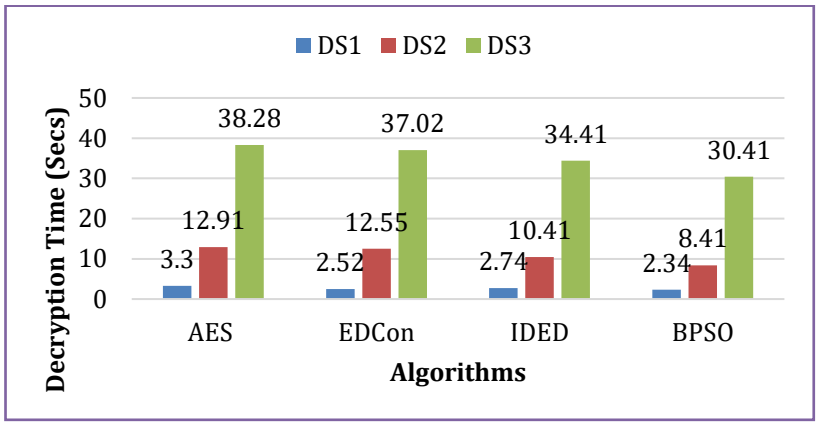

Fig 3a. Comparison of results w.r.t. decryption time on small size datasets

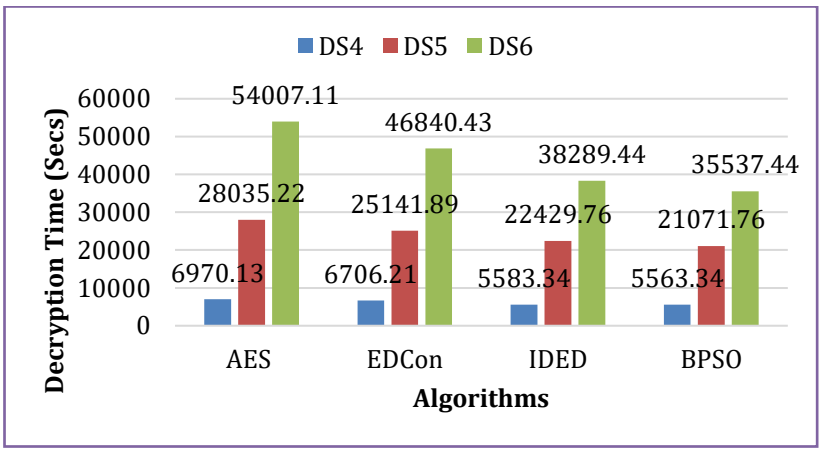

Fig 3b. Comparison of results w.r.t. decryption time on large size datasets

On comparing the results of 3a, on small sized datasets the proposed model outperforms significantly when compared with existing methods w.r.t. decryption time. In specific, on DS1, BPSO outperforms AES with 29.1%, EDCon with 7.1% and IDED with 14.6%. On DS2, BPSO outperforms AES with 34.9%, EDCon with 33% and IDED with 19.2%. On DS3, BPSO outperforms AES with 20.6%, EDCon with 17.9% and IDED with 11.6%. On comparing the results in Figure 3b, in large datasets DS4 BPSO outperforms AES with 20.2%, EDCon with 17% and IDED with 0.4%. On DS5, BPSO outperforms AES with 24.8%, EDCon with 16.2% and IDED with 6.1%. On DS6, BPSO outperforms AES with 34.2%, EDCon with 24.1% and IDED with 7.2%.

5.2.3 *Experimental Evaluation w.r.t memory usage*

Table 4. Experimental Results of proposed Vs existing algorithms w.r.t memory usage

DataSet	AES	ICA	IDED	BPSO
DS1	38925	37688	35885	35111
DS2	43406	38020	37868	37866
DS3	213420	186681	185493	185489
DS4	269762	250282	234350	234319
DS5	623988	607820	582205	580898
DS6	940502	919561	892783	889956

Table 4 shows the memory usage in taken by the proposed and existing algorithm on large and small datasets. Figure 4a and 4b shows the graphical representation of results in Table 4 for better interpretation.

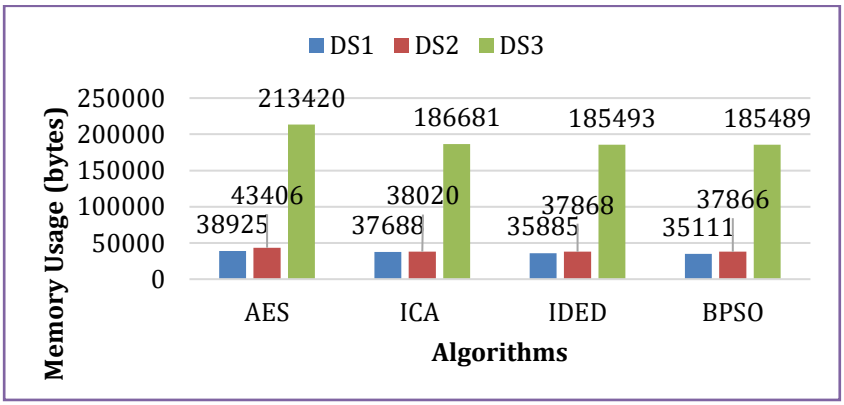

Fig 4a. Comparison of results w.r.t. memory usage on small size datasets

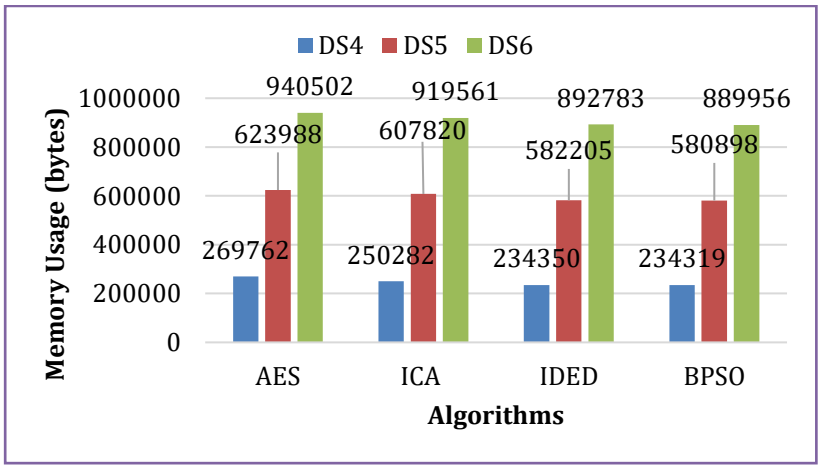

Fig 4b. Comparison of results w.r.t. memory usage on large size datasets

On comparing the results of 4a, on small sized datasets the proposed model outperforms significantly when compared with existing methods w.r.t. memory usage. In specific, on DS1, BPSO outperforms AES with 9.8%, ICA with 6.8% and IDED with 2.2%. On DS2, BPSO outperforms AES with 12.8%, ICA with 0.4% and performs equally with IDED. On DS3, BPSO outperforms AES with 13.1%, ICA with 0.6% % and performs equally with IDED. On comparing the results in Figure 3b, in large datasets DS4 BPSO outperforms AES with 13.1%, ICA with 6.4% and performs equally with IDED. On DS5, BPSO outperforms AES with 6.9%, ICA with 4.4% and IDED with 0.2%. On DS6, BPSO outperforms AES with 5.4%, ICA with 3.2% and IDED with 0.3%.

6. Conclusion

Our suggested architecture is an intelligent and efficient secrecy maintaining framework on distributed cloud storage, as shown by the detailed comparison study presented before. To protect the privacy of the ever-

increasing amounts of data kept in the cloud, it may serve as a filter. When used as a filter, the framework will not only use less processing time, but also much less memory. As large data poses unique challenges for managing both time and memory, we recommend adopting the suggested framework to protect the privacy of data being moved to the cloud. Our suggested architecture may be further developed such that methods like deep learning (using the dropout technique) and fuzzy logic can be employed to generate Crypto-Nets, and these in turn can be utilised to form safe and intelligent cloud storage units that can detect confidentiality breaches.

References

[1] Walker, SH; Duncan, DB (1967). "Estimation of the probability of an event as a function of several independent variables". Biometrika. **54** (1/2): 167–178

[2] Y. Li, K. Gai, L. Qiu, M. Qiu, H. Zhao, Intelligent cryptography approach for secure distributed big data storage in cloud computing, Inform. Sci. 387 (2017) 103–115

[3] N. Cao, C. Wang, M. Li, K. Ren, W. Lou, Privacy-preserving multi-keyword ranked search over encrypted cloud data, IEEE Trans. Parallel Distrib. Syst. 25 (1) (2013) 222–233.

[4] Y. Li, K. Gai, L. Qiu, M. Qiu, H. Zhao, Intelligent cryptography approach for secure distributed big data storage in cloud computing, Inform. Sci. 387 (2017) 103–115.

[5] Kamal, Maryam, Shahzad Amin, Faria Ferooz, Mazhar Javed Awan, Mazin Abed Mohammed, Omar Al-Boridi, and Karrar Hameed Abdulkareem. "Privacy-aware genetic algorithm based data security framework for distributed cloud storage." Microprocessors and Microsystems 94 (2022): 104673.

[6] A. Mehmood, I. Natgunanathan, Y. Xiang, G. Hua, S. Guo, Protection of big data privacy, IEEE Access 4 (2016) 1821–1834.

[7] S. Han, K. Han, S. Zhang, A data sharing protocol to minimize security and privacy risks of cloud storage in big data era, IEEE Access 7 (2019) 60290–60298.

[8] S.M. Idrees, M.A. Alam, P. Agarwal, A study of big data and its challenges, Int. J Inform. Technol. 11 (4) (2019) 841–846.

[9] M. Strohbach, J. Daubert, H. Ravkin, M. Lischka, Big data storage. In New horizons for a data-driven economy (pp. 119-141). Springer, Cham.Author 1, A.; Author 2, B. Title of the chapter. Book Title, 2nd ed., Publisher, Publisher Location, Country, 2016, pp. 154–196. Editor 1, A., Editor 2, B., Eds.2007; Volume 3.

[10] S. Wang, Y. Zhang, Y. Zhang, A blockchain-based framework for data sharing with fine-grained access control in decentralized storage systems, IEEE Access 6 (2018) 38437–38450

[11] S. Singh, J. Sidhu, Compliance-based multi-dimensional trust evaluation system for determining trustworthiness of cloud service providers, Fut.Gener. Comput. Syst. 67 (2017) 109–132

[12] R. Nachiappan, B. Javadi, R.N. Calheiros, K.M. Matawie, Cloud storage reliability for big data applications: A state of the art survey, J. Netw. Comput. Appl. 97 (2017) 35–47.

[13] G. Baranwal, D.P. Vidyarthi, A framework for selection of best cloud service provider using ranked voting method, in: 2014 IEEE international advance computing conference (IACC), IEEE, 2014, pp. 831–837.

[14] L. Zhang, H. Xiong, Q. Huang, J. Li, K.K.R. Choo, L.I. Jiangtao, Cryptographic solutions for cloud storage: Challenges and research opportunities, IEEE Trans. Serv. Comput. (2019).

[15] R. Fardel, M. Nagel, F. Nüesch, T. Lippert, A. Wokaun, Fabrication of organic lightemitting diode pixels by laser-assisted forward transfer. Appl. Phys. Lett. 91 (6) (2007), 061103.

[16] S. Balasubramaniam, V. Kavitha, A survey on data encryption tecniques in cloud computing, Asian J. Inform. Technol. 13 (9) (2014) 494–505.

[17] P Velmurugadass, S Dhanasekaran, SS Anand, V. Vasudevan, Enhancing Blockchain security in cloud computing with IoT environment using ECIES and cryptography hash algorithm, Mater. Today: Proc. 37 (2021 Jan 1) 2653–2659, https://doi.org/10.1016/j.matpr.2020.08.519.

[18] F. Thabit, A.P.S. Alhomdy, A.H. Al-Ahdal, S. Jagtap, A new lightweight cryptographic algorithm for enhancing data security in cloud computing, in: Global Transitions Proceedings, 2021, https://doi.org/10.1016/j.gltp.2021.01.013.

[19] K. Samriya, N. Kumar, A novel intrusion detection system using hybrid clustering-optimization approach in cloud computing, Mater. Today: Proc. (2020), https://doi.org/10.1016/j.matpr.2020.09.614.

[20] A. Ullah, S.M. Siddiquee, M.A. Hossain, S.K. Ray, An ethereum blockchain-based prototype for data security of regulated electricity market, Inventions 5 (4) (2020) 58.

[21] R. Ngnie Sighom, P. Zhang, L You, Security enhancement for data migration in the cloud, Future Internet 9 (3) (2017) 23.

[22] X. Zhang, C. Liu, S. Poslad, K.K. Chai, A provable semi-outsourcing privacy preserving scheme for data transmission from IoT devices, IEEE Access 7 (2019) 87169–87177.

[23] Q. Xu, C. Tan, Z. Fan, W. Zhu, Y. Xiao, F. Cheng, Secure multi-authority data access control scheme in cloud storage system based on attribute-based signcryption, IEEE Access 6 (2018) 34051–34074.

[24] J.K. Liu, K. Liang, W. Susilo, J. Liu, Y. Xiang, Two-factor data security protection mechanism for cloud storage system, IEEE Trans. Comput. 65 (6) (2015) 1992–2004.

[25] Thirugnanasambandam, K.; Ramalingam, R.; Mohan, D.; Rashid, M.; Juneja, K.; Alshamrani, S.S. Patron–Prophet Artificial Bee Colony Approach for Solving Numerical Continuous Optimization Problems. Axioms **2022**, 11, 523. https://doi.org/10.3390/axioms11100523

[26] Thirugnanasambandam, K., Rajeswari, M., Bhattacharyya, D. et al. Directed Artificial Bee Colony

algorithm with revamped search strategy to solve global numerical optimization problems. Autom Softw Eng **29,** 13 (2022). https://doi.org/10.1007/s10515-021-00306-w

[27] Raghav RS, Thirugnanasambandam K, Varadarajan V, Vairavasundaram S, Ravi L. Artificial Bee Colony Reinforced Extended Kalman Filter Localization Algorithm in Internet of Things with Big Data Blending Technique for Finding the Accurate Position of Reference Nodes. Big Data. 2021 Nov 5. doi: 10.1089/big.2020.0203. Epub ahead of print. PMID: 34747652

[28] Thirugnanasambandam, K., Raghav, R.S., Anguraj, D.K. et al. Multi-objective Binary Reinforced Cuckoo Search Algorithm for Solving Connected Coverage target based WSN with Critical Targets. Wireless Pers Commun (2021). https://doi.org/10.1007/s11277-021-08824-2

[29] Thirugnansambandam, K., Bhattacharyya, D., Frnda, J., Anguraj, D. K., Nedoma, J. (2021). Augmented Node Placement Model in t-WSN Through Multiobjective Approach. CMC-Computers, Materials & Continua, 69(3), 3629–3644

[30] Thirugnanasambandam, K., Anitha, R., Enireddy, V. et al. Pattern mining technique derived ant colony optimization for document information retrieval. J Ambient Intell Human Comput (2021). tps://doi.org/10.1007/s12652-020-02760-y