Frequent Pattern Mining (FPM) - Privacy Preserving and Security of Intermediate Data in Cloud Storage

V. Sarala, Research Scholar
Department of Computer science and Engineering, SCSVMV University, Kanchipuram, India
Dr. C. K. Gomathy, Assistant Professor
Department of Computer science and Engineering, SCSVMV University, Kanchipuram, India

Abstract
Nowadays, cloud computing has played a vital role in most data-intensive applications to store the data in the intermediated dataset. This effective cloud storage process helps to minimize the storage and processing cost while performing recomputing. Although the cloud provides numerous services, resources maintaining cost, outsourced user data protection from unauthorized users, the privacy of sensitive data, and computation complexity is still a major issue. A novel network-based frequent pattern mining model (DLFPM) is introduced to overcome these issues. Here, the presented method examines the frequent access information according to the layers of network functions. The network computes the sensitive information from the deep learning function. The identified sensitivity information is encrypted using the single sign-on associated with the Paillier encryption technique (SSO-PE) that avoids unauthorized access. The effective utilization of these algorithms continuously manages the sensitive data security that helps to minimize the computation cost and computation time.
Keywords*:*Cloud computing, intermediate dataset, unauthorized access, novel deep learning network based frequent pattern mining model, sign-on associated with Paillier encryption technique

1. Introduction

Cloud Storage [1] is an important part of the cloud computing model that helps to save data and valuable information via the internet. The cloud provider gives the cloud storage [2,3]as services to the user based on their demands with minimum cost and effectively manages the data storage infrastructure. During this process, users store the intermediate dataset [4] information of specific intensive applications to minimize the frequent re-computation of the original dataset. The users can reanalyze the results, share intermediate results, and conduct new analysis to other for making effective collaborations[5-7]. The shared intermediate data accessed by intermediate users, third-party and unauthorized users, causes to create the security issues. Therefore, the shared data encrypted and stored in the cloud storage for minimizing unauthorized activities. This security process consumes more time and makes the computational complexity.

The other challenging task is, most of the application uses the unencrypted data [8] to process their information. However, the user wants to manage the data security and privacy [9,10] while accessing and storing their data in the cloud environment. To achieve this, researchers use various encryption techniques [11,12] such as Attributed-based Encryption (ABE), Advanced Encryption Standard (AES), Data Encryption Standard (DES), Elliptic Curve Cryptography (ECC) approach, RSA, etc. for encrypting data. Although traditional encryption techniques work well, encrypting the entire dataset is complex and time-consuming. Therefore, researchers are doing their research to identify the intermediate datasets and encrypt the data using encryption techniques [13-15] to avoid the discussed issues. This type of encryption process minimizes the computation cost while persevering privacy in the cloud. However, identifying intermediate datasets and sensitive information is a complex task while providing privacy to the data. This is achieved by applying the datamining techniques [16,17] like support vector machine, frequent data analysis, matching rules, and neural networks. These algorithms successfully recognize the sensitive information and intermediate dataset using frequent data access and learning rules. Once the information is identified, it is a very easy task to encrypt the data.

As discussed, various encryption techniques are presented to maintain data security and privacy. Nevertheless, the existing approaches are weak security frameworks that fail to restrict the unauthorizedaccess while outsourcing data in the cloud environment. Several intermediate attacks and third-party access is difficult to recognize during the data accessing and storing process. In addition to this, traditional approaches consume more time and cost to manage privacy preserving for high-dimensional datasets. A novel deep learning network-based frequent pattern mining model (DLFPM) is introduced to resolve these research issues in this work. The DLPM approach uses the learning process and function in every layer, which helps predict frequent access to information from the dataset. More ever, it able to identify sensitive information from the trained network. The identified sensitivity information is encrypted using the single sign-on associated with the Paillier encryption technique that effectively avoids unauthorized access. This discussed system implemented using the .Net framework, and the system's effectiveness is examined using Computation cost, complexity, and security metrics. According to the discussion, the overall objective of the work is summarized as follows.

- To minimize the sensitive information and stored data maintaining cost in the cloud.
- To reducing the unauthorized access to outsourced user data in the cloud.
- Providing guarantee for data protection, reducing third-party involvement, avoiding recurrence attacks by implementing a strong security framework.
- Ensure security and privacy for data owners while saving data in the cloud with minimum computation complexity.

Then the rest of the paper is organized as follows; section 2 discusses several researchers' works on privacy preservation on cloud environment data storage. Section 3 discusses the detailed working process of a novel deep learning network-based frequent pattern mining model and single sign-on associated with the Paillier encryption technique (DLFPM-SSO-PE). The efficiency of the DLFPM-SSO-PE approach is discussed in section 4, and the conclusion is discussed in section 5.

2. Related Works

C. Xu, et al., 2019 [18] developing secure and privacy-based health data sharing in cloud systems. This process uses a searchable encryption technique (PPSE) which uses the multi-keyword searching process to manage data privacy. In addition to this, bloom filter and message authentication code investigate personal health information (PHI). This investigation process manages the data integrity and filters the false data successfully.

S. Sharma et al.[19] introduced a private graph to make the spectral analysis in the cloud. The spectral analysis consists of data contributors, data owners, and cloud providers while utilizing the privacy-preserving algorithm. The defined matrices are definedin terms of adjacency and Laplacian matrix, encrypted and replaced by the distributed contributors. Further, data privacy is ensured using Lanczos and Nystrom algorithm. These algorithms use somewhat homomorphic encryption and additive homomorphic encryption. The defined algorithms ensure the minimum cost and computation time compared to other methods.

A. Yang et al.,[20] storing data in cloud environments using lightweight and privacy-preserving delegatable proofs. The proof of storage process ensures data privacy on one side, and close functionalities help audit third-party access on other sides. Therefore, the delegatable assuranceprovides security at any time with minimum time computations and attains high efficiency while storing data in the cloud environment.

Singh, N., et al., [21] reviewing various privacy-preserving methodologies to manage security to the stored data in the cloud. This analysisis categorized according to privacy by probability, privacy by cryptography, privacy by ranking, and privacy by anonymization. This categorization helps to analyzing and auditing the data against unauthorized data access. Thus, the discussed methodologies ensure data confidentiality and authentication successfully.

Dave J. et al.,[22] applying the deduplication process to minimize the data storage cost. This process reduces the data size, and the malicious user activities are eliminated with the help of a secure

random key encryption technique. This process encrypts the files using the random key, and the hash values are utilized to generate the plaintext. The random key generation-based encryption helps to improve data security and reduce dictionary and brute-force attacks effectively.

D. N. Wu, et al.[23] developed a verifiable public key homomorphic encryption technique to maintain data security and privacy in cloud storage. During the encryption process, the server creates the inverted encryption index structure that helps to enhance the searching efficiency. In addition to this, the authentication structure is maintained by enabling multi-user settings that verify the user's correctness and completeness. Then the discussed system overcomes the approximate GCD problem in a cloud storage with minimum computation complexity.

C. Guo et al.,[24] creating the secure cloud-IoT Ecosystems using the K-Nearest Neighbor (KNN) Query over encryption process (KNN-HE). The KNN algorithm investigates the uncertain data presented in the semi-trusted cloud server. The retrieved data is encrypted by applying the modified homomorphic encryption technique. After that authorized rank method is used to estimate the KNN related uncertaindata. This KNNquery-based encryption process improves overall data security and privacy with minimum time.

A. A. Badawi et al., [25]performing fast and private text classification using fully homomorphic encryption technique (PrivFT). Initially, the plaintext model is generated by encrypting user inputs and the training model to the encrypted data. This training process minimizes the loss value and improves the classification performance. The discussed system utilizes the natural language processing public dataset to perform these tasks with 0.17s time.

Mohanty et al. [26] providing IoT Security and privacy to data by applying a lightweight, integrated blockchain approach. This model is implemented in the smart home environment in which the captured data is transferred from one location to another by providing data security. Here, three optimization algorithms, such as certificateless cryptography, lightweight consensus algorithm, and distributed throughput management scheme,ensure data security.

Kaaniche et al. [27] applying cryptographic techniques to provide data security and privacy while storing data in a cloud storage environment. The system intends to resolve the data control issues, confidentiality, and privacy issues using various cryptographic algorithms.

Shah et al.,[28] developing blockchain-based decentralized cloud storage system. Here, user files are encrypted in the multiple peer network, using interplanetary file system protocol. The protocol uses the hash function to identify the file path and the encrypted file stored according to the blockchain process.

Riad et al. [29] creating an IoT-based multi-authority-based cloud storage system by providing hierarchical access control. The introduced methods were analyzing the users and providing multiple authorities to access the data. This system encrypts the data in the cloud system and ensures privacy, security, and confidentiality by eliminating attacks.

According to the various researchers' opinions, data security and privacy are more important while storing data in a cloud environment. The previous study presents several encryption techniques, but they fail to maintain a strong security framework while implementing the security to the data in the cloud. Therefore, an effective and secure system is developed to manage the cloud's intermediate data storage privacy in this work. The detailed working process is illustrated as follows that helps to resolve the discussed research problems.

3.DLFPM-SSOPE-based secure data storage in the cloud

This section discusses the deep learning network-based frequent pattern mining model and single sign-on associated with Paillier encryption technique (DLFPM-SSO-PE) based secure data storage process in the cloud environment. Here, the intermediate data is stored in the cloud for making further research analysis. During this process, intermediate users try to access the data that causes unauthorized access and reduces privacy and security. The stored information is more sensitive, which needs to be protected against unauthorized access by applying the DLFPM approach. The introduced method predicts the sensitive information from the intermediate dataset according to the

frequent access patterns. The expected information is encrypted according to the single sign-on associated with the Paillier encryptionmethod and saved in the encrypted format. This process helps to improve the overall data security and privacy while storing data in the cloud environment.

DLFPM system architecture. Initially, the data owners upload their documents and files in cloud storage to make future research analyses. Generally, the saved data is huge in dimension,consuming more processing time and cost while planning to provide security. Therefore, the data owners or users process the dataset according to their requirements and create intermediate data to make their analysis easier. When the users are training to access the intermediate dataset by authenticating the user details, which helps to avoid unauthorized access. The stored intermediate data is saved in the encrypted format, which is performed according to the frequent data access.

First, frequent data must be analyzed in the intermediate dataset by applying the frequent pattern mining algorithm. Here Frequent Pattern (FP) algorithm is utilized to examine the simple structure, frequent patterns, and data association in the dataset. The frequent mining process identifies the items-related transactions in the database.

By considering, the frequent patterns are extracted according to the support (S), confidence (C), and lifting (L)rules. Consider the dataset has n attributes or items that are represented as $I = \{i_1, i_2, i_3 \ldots\ldots i_n\}$. For every attribute has the respective transactions, which is denoted as $D = \{t_1, t_2, t_3 \ldots\ldots t_n\}$. These transactions havea unique transaction ID that is relevant to I. As discussed, the frequent pattern is identified by three rules (S,C, and L). The Support (S) rule is used to investigate how frequently the data or attribute appears in the dataset. Considered, X and Y areitem sets defined by association rule $X \rightarrow Y$, and the given dataset transaction is defined as T. Then the frequently appeared data is identified by eqn (1) based rule.

$$S(X) = \frac{|X \subseteq T|}{|T|} \quad (1)$$

The frequently appeared data $S(X)$ is identified according to the proportion of transaction T in the database contains the X itemset. After identifying the frequently appeared data, how often the recognized value is true should be investigated that is done by using the C rule. The C rule on $X \rightarrow Y$ is computed using eqn (2).

$$C(X \rightarrow Y) = \frac{S(X \cup Y)}{S(X)} \quad (2)$$

The C value is computed with the help of conditional probability, namely, $P(E_Y|E_X)$; E_X and $E_Y$is the event thsat transactions consist of X and Y itemset. This process helps to identify the number of times the frequent data appeared in the dataset. At last, lifting rule (L) should be applied according to eqn (3)

$$L(X \rightarrow Y) = \frac{S(X \cup Y)}{S(X) * S(Y)} \quad (3)$$

The computed L rule used to identify the attributes dependency$S(X \cup Y)$ and independency $(S(X)\ and\ S(Y))$ in the dataset. By considering these rules, the frequent pattern is examined according to the user-specified minimum support and confidence value at the same time. These two rules are placed a vital role in selecting the frequent items in the dataset. Once the frequent items are chosen, which are considered as the features and transmitted to the deep learning concept to identify the sensitive information. This work uses the long short-term memory deep neural network (LSTM) to perform the sensitive information identification from the extracted frequent pattern. The LSTM network helps resolve the long-term dependency problem by memorizing the data for a long time. The network uses the memory cells that are used to track the data dependence throughout the dataset. The extracted frequent patterns are considered as input which is processed by input gate, output gate, forget gate, and update gate. These layers of gates help to identify the data dependency and the output prediction process.

The deep learning network utilizes its activation layers and gates function to predict the sensitive information from the frequent pattern inputs. The activation layer is presented in the activation layer used to identify the output for the respective input. According to the sigmoid function, the memory cell continuously tracks the data dependency and updates in the update gate. The incoming inputs are fed into the forget gate to decide whether the memory cell should forget or leave the data dependency. Based on the discussion, the activation layer tanH value is identified by using the input layer weight (W_c), bias (b_c) and previous activation layer value ($a^{(t-1)}$).

$$\tilde{c}^{(t)} = tanh\left(W_c\left[a^{(t-1)}, x^{(t)}\right] + b_c\right) \quad (4)$$

The computation of $\tilde{c}^{(t)}$ in tanh gives the value between -1 and 1, which helps identify the dependency of input (frequent pattern) and sensitive information. Once the dependence is determined, the decision is made, the tracked values are stored or not in the memory cell. This is done by using the forget and update gate (sigmoid layers).

$$\Gamma_u = \sigma\left(W_u\left[a^{(t-1)}, x^{(t)}\right] + b_u\right) \quad (5)$$

$$\Gamma_f = \sigma\left(W_f\left[a^{(t-1)}, x^{(t)}\right] + b_f\right) \quad (6)$$

According to the computation, the values are updated in the memory cell using the sigmoid layer function that is defined in eqn (7)

$$\Gamma_o = \sigma\left(W_o\left[a^{(t-1)}, x^{(t)}\right] + b_o\right) \quad (7)$$

$$c^{(t)} = \Gamma_u * \tilde{c}^{(t)} + \Gamma_f * c^{(t-1)} \quad (8)$$

After updating the computed values, the output is calculated using the SoftMax function; if it returns one, the frequent pattern belongs to sensitive information. The activation function in the network is computed using eqn (9).

$$a^{(t)} = \Gamma_o * \tanh\left(c^{(t)}\right) \quad (9)$$

This process is implemented by using a two-layer of LSTM model with 512 hidden nodes in every layer. The memory cell utilized in the network helps to identify the sensitive information with minimum time. Also,the network updates its memory continuously that used to improve the overall frequent pattern-related sensitivity. Once the information is identified from the intermediate dataset, it has been stored in the cloud to make other analyzing and research purposes. Data should be kept back in an encrypted format to reduce unauthorized access and third-party attacks.

Therefore, in this work, data saved by encrypting the single sign-on (SSO) associated with the Paillier encryption technique. The SSO process identifies the users by using other trusted sites, which is the initial verification process of the user while accessing the data. In otherwords, it is saying the authentication process that maintains the trust relationship between the user and service provider in the cloud. The trust is supported by exchanging the certification between the service provider and identity provider. After completing the initial verification process of the user, the data is encrypted using the Paillier encryption technique. It is one of the probabilistic asymmetric public-key cryptography algorithms and additive homomorphic cryptosystem. The algorithm works by ensuring three steps such as key generation, encryption, and decryption. The detailed encryption process is illustrated in figure 1.

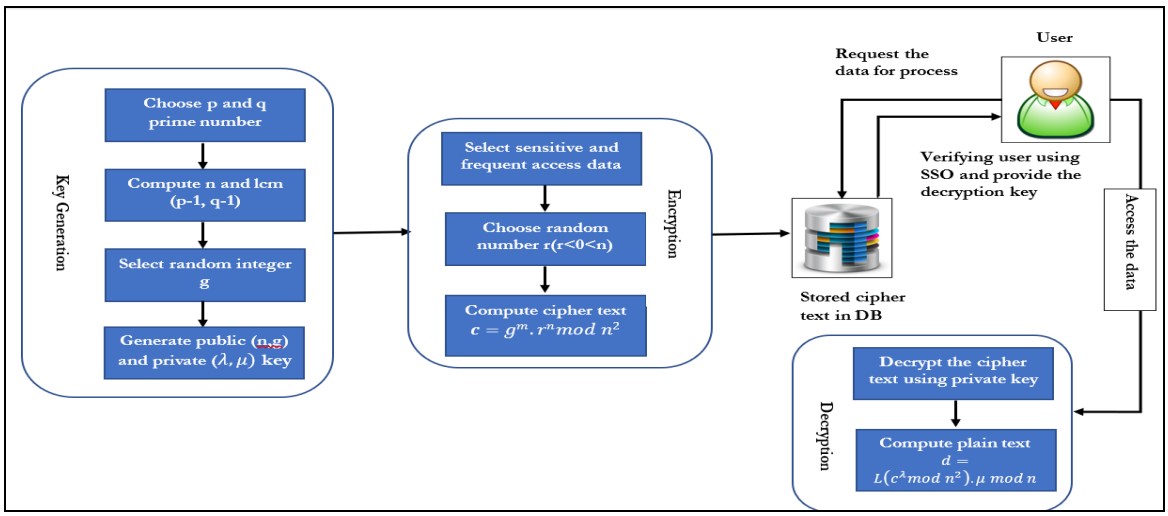


Figure 1: Process of Security establishment using Paillier Encryption technique

Key generation

Initially, the key has to be generated for encrypting the shared data in the cloud. For that, large prime numbers p and q is selected randomly and independently of each other like $\gcd(pq,(p-1)(q-1))=1$. Here, $n=pq$ and $\lambda=lcm(p-1,q-1)$. Then random integer g must be chosen$g\in\mathbb{Z}_{n^2}^*$. Then compute the modular multiplicative inverse value $\mu=\left(L(g^\lambda mod\ n^2)\right)^{-1}\ mod\ n$. Here L is estimated as$L(x)=x-1/n$. According to the defined values, the public and private keysare generated as follows.

$$public\ key=(n,g)\ and\ private\ key=(\lambda,\mu) \quad (10)$$

Suppose pand q are the same lengths, then the key-value parameter computation changed slightly. Where $g=n+1, \lambda=\varphi(n), \mu=\varphi(n)^{-1}$ mod n ; $\varphi(n)=(p-1)(q-1)$.

Encryption

After generating the key value, data m should be encrypted using the private and public keys. During the encryption, process data is $0\le m<n$. Then random integer r $0<r<n; and\ r=\mathbb{Z}_{n^2}^*$.is selected for making the encryption process. Then the data ciphertext is obtained according to eqn (11)

$$cipher\ text\ c=g^m.r^n mod\ n^2 \quad (11)$$

The encrypted data is kept in the cloud database. Here, data is accessed only after authenticating user details. The introduced method uses the SSO process, which requires user credential information that is difficult to guess by intermediate users. In addition to this, the key values are selected randomly in an earlier stage which also minimizes the intermediate access. Once the user verification is completed decryption key is given to the user for accessing the data.

Decryption

Here, the ciphertext c has to be decrypted, and the c value belongs to $\mathbb{Z}_{n^2}^*$. By using the decryption key, the plain text d is obtained as follows.

$$d=L(c^\lambda mod\ n^2).\mu\ mod\ n \quad (12)$$

The Paillier algorithm uses the homomorphic properties during the encryption and decryption process while performing addition and multiplication parameters. The effective utilization of these parameters manages data security because there is no way to compute the plain text without expressive the

private key. Thus, the introduced system eliminates the existing research issues by implementing the frequent pattern-based encryption algorithm.

4. Results and Discussion

The discussed deep learning network-based frequent pattern mining model and single sign-on associated with Paillier encryption technique -data security and privacy-preserving system efficiency- is evaluated. This system computes the association between the data attributes and respective transactions, which helps predict the dataset's frequent patterns. The continuous prediction of frequent patterns minimizes the computation complexity also the redundancy issue of the encryption process. Then deep learning approach is applied to classify the frequent pattern characteristics used to select the sensitive information. This process avoids the difficulties involved in the sensitive information management process. Instead of providing the entire dataset, this DLFPM process identifies the most frequent sensitive information that minimizes the computation cost compared to the other methods. In addition to this, the encryption process authenticates user information before accessing and storing the data. This authentication eliminates the unauthorized user's involvement in cloud data storage and access. It minimizes the computation cost and complexity while accessing the intermediate data. Then the discussed system efficiency is evaluated using the computation cost, computation time, and accuracy of security..

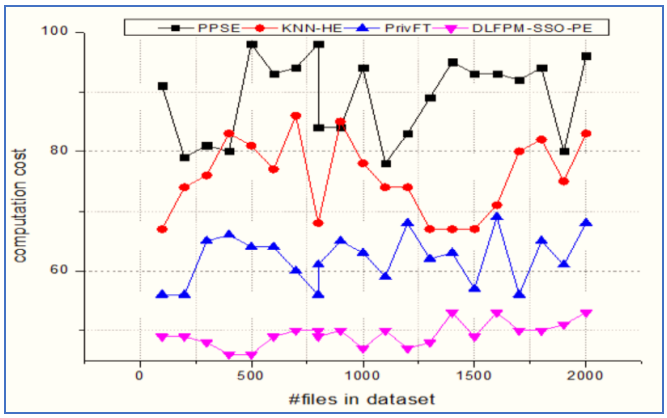


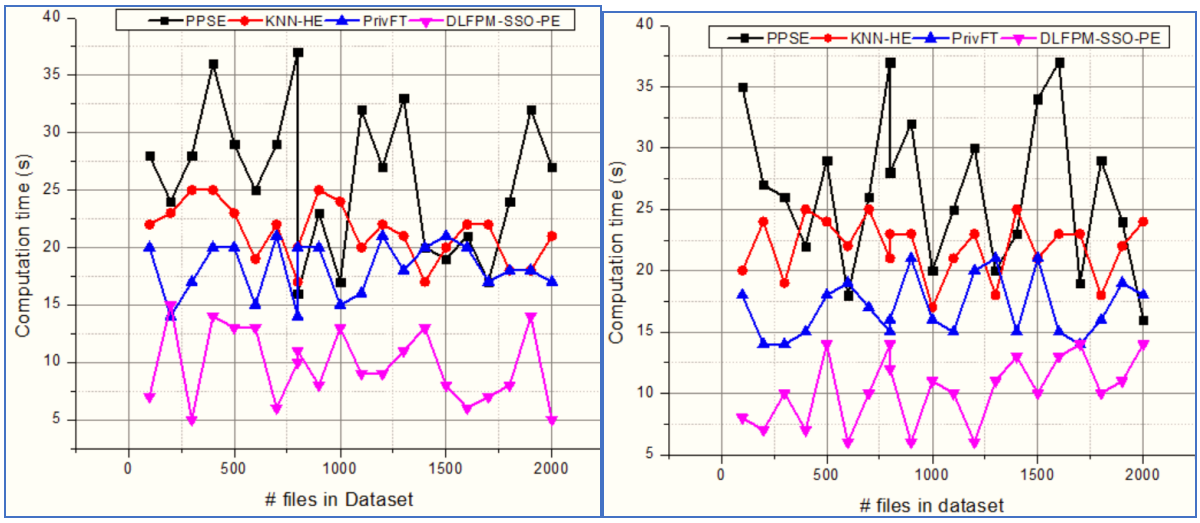
Figure 2: Computation cost

Figure 2 illustrated that the computation cost for encrypting the sensitive information in the original dataset. As discussed earlier, the method consumes minimum cost for analyzing the frequent pattern, sensitive information classification, and encryption process. The introduced method examines the attributes dependency $S(X \cup Y)$ and independency ($S(X)$ and $S(Y)$) in the dataset according to S, C, and L. Then, the deep learning network uses the memory cell, which updates every frequent pattern in memory. This reduces the computation cost for analyzing the entire dataset attributes. Then the encryption and decryption process uses the effective key $public\ key = (n, g)\ and\ private\ key = (\lambda, \mu)$Which also minimizes the computation cost for the different number of files in the dataset. In addition to this, the system consumes minimum computation time to perform the frequent data analysis, sensitive data classification, and security establishment process. Then the obtained computation cost for frequent and sensitive information is illustrated in table 1.

Table 1: Computation time

	Frequent data access computation time									
Number of files	**200**	**400**	**600**	**800**	**1000**	**1200**	**1400**	**1600**	**1800**	**2000**
PPSE	24.33	36.2	25.2	16.87	17.39	27.2	20.19	21.18	24.19	27.19
KNN-HE	23.79	25.19	19.2	20.15	24.39	22.16	17.29	22.37	18.37	21.38
PrivFT	14.2	20.38	15.16	20.97	15.1	21.3	20.19	20.54	18.18	17.27
DLFPM-SSO-PE	15.28	14.29	13.83	11.19	13.83	9.98	13.23	6.29	8.23	5.48
Number of files	**Sensitive data identification computation time**									
PPSE	24.40	18.93	26.34	23.95	23.69	34.78	29.11	32.88	27.46	20.534
KNN-HE	17.07	22.74	18.27	23.10	19.85	24.188	23.58	19.30	22.99	21.40
PrivFT	18.63	18.18	16.47	18.04	14.57	18.86	17.20	19.59	18.47	19.56
DLFPM-SSO-PE	8.80	8.99	9.62	9.45	10.54	13.59	7.798	13.62	7.720	12.46

From the table 1 it clearly shows that introduced DLFPM-SSO-PE method attains minimum computation time for accessing frequent data and sensitive information from intermediate dataset.



(a) (b)

Figure 3 :(a) Frequent pattern and sensitive information identification and (b) encryption and decryption

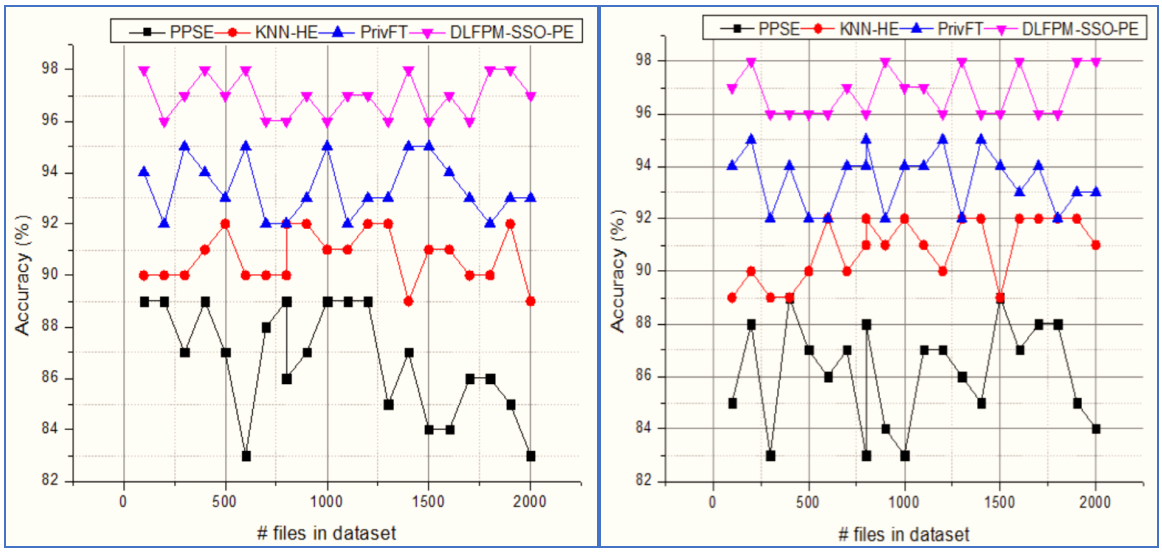
The figure 3 (a) and (b) illustrated that the encryption time of entire system working process. In figure (a), the introduced method attains minimum computation time compared to other researcher's work C. Xu, et al., 2019 [18], C. Guo, et al.,[24] and A. A. Badawi, et al., [25]. Here, the frequent patterns are examined according to the association rules S ($(X \to Y) = \frac{S(X \cup Y)}{S(X)}$), C($S(X) = \frac{|X \subseteq T|}{|T|}$) and L($L(X \to Y) = \frac{S(X \cup Y)}{S(X) * S(Y)}$). This helps to compute the upcoming data and respective frequent patterns with minimum time. Further, the system is trained according to the frequent pattern using the deep learning approach. The network uses the memory cell for updating every frequent pattern and respective sensitive information computed by the output gate Γ_o and $c^{(t)}$. This process completely minimizes the computation time. After that, a single sign-on (SSO) algorithm validated the users and authorized the user in the initial stage itself; this helps improve the system's overall security. The initial stage of authentication also minimizes the additional processing time. The encryption and decryption

keys quickly encrypt the user data, and the obtained result is shown in figure (b). Then the overall accuracy of the system is illustrated in table 2.

Table 2: Accuracy

	Frequent data access accuracy									
number of files	200	400	600	800	1000	1200	1400	1600	1800	2000
PPSE	88.36	88.49	87.38	88.05	88.58	87.72	88.28	88.22	87.72	87.56
KNN-HE	90.54	89.06	90.27	89.12	90.42	91.78	90.21	90.54	90.42	89.33
PrivFT	94.4	92.44	92.95	93.08	93.16	93.73	92.91	94.68	92.77	92.17
DLFPM-SSO-PE	97.38	97.12	97.95	97.59	97.36	97.82	97.48	97.29	97.5	97.55
number of files	**Sensitive data identification accuracy**									
PPSE	86.28	85.49	85.14	87.44	84.56	88.5	88.88	88.55	88.87	88.11
KNN-HE	93.67	94.48	92.64	92.65	94.15	93.81	92.05	94.67	93.34	94.83
PrivFT	94.47	94.25	94.28	94.03	92.81	93.45	92.7	93.35	93.51	93.43
DLFPM-SSO-PE	96.48	96.9	97.24	97.82	96.7	96.43	96.35	97.4	97.61	96.19

According to table 2, the obtained accuracy value related graphical representation is shown in figure 4.



(a) (b)

Figure 4: Accuracy (a) Frequent pattern and sensitive information identification and (b) encryption and decryption

The figure 4 (a) and (b) illustrated that the accuracy of entire system working process. In figure (a), the introduced method attains minimum computation time compared to other researcher's work C. Xu, et al., 2019 [18], C. Guo, et al.,[24] and A. A. Badawi, et al., [25]. Here, the frequent patterns are examined according to the association rules,and the data dependency helps to identify the correct data with maximum accuracy.According to the frequent pattern computation, the upcoming future information was also identified correctly. Further, the system is trained according to the frequent pattern using long short term memory deep learning network. The network uses the memory cell for updating every frequent pattern and respective sensitive information computed by the output gate

Γ_o and $c^{(t)}$. This process adjusts the network weight and bias values according to the training function. This causes to minimizes deviations and improves the overall security rate. In addition to this, the system uses the SSO algorithm to perform the initial authentication process, which almost provides the highest security level. However, the system uses random encryption and decryption keywords to perform the encryption and decryption process. The effective computation of attributes conditional probability values and respective frequent patterns and encrypting keys are difficult to guess by intermediate users. This leads to improving the overall security of the system.

In addition to this, the proportion execution time for processing plaintext in the intermediate dataset. Here, the Paillier encryption algorithm generate the private and public keys for processing the data and convert into the unreadable format. This helps to improve the overall data security also the generated cipher text and respective key values are difficult to access by third-party access. The plaintext into ciphertext conversion process should be consume minimum execution time. Then the obtained proportion execution time for plaintext processing is illustrated in table 3.

Table 3: Execution time for plaintext processing

	Proportion execution time for plaintext processing (s)									
Number of files	**200**	**400**	**600**	**800**	**1000**	**1200**	**1400**	**1600**	**1800**	**2000**
PPSE	14.49	14.39	15.36	14.23	13.54	17.71	14.05	16.3	15.34	13.92
KNN-HE	9.96	10.01	11.09	10.28	10.61	11.97	10.73	9.44	11.97	11.21
PrivFT	9.81	10.52	9.34	10.23	8.42	9.66	8.15	8.78	10.04	8.77
DLFPM-SSO-PE	3.33	3.23	2.28	3.42	2.87	3.2	2.04	3.1	2.36	3.89

From the table 3 it clearly shows that introduced DLFPM-SSO-PE method consumes minimum execution time for converting the plaintext into the cipher text. This directly indicates the SSO-PE approach ensures the data security while storing in the third-party server. The obtained results are very low compared to existing methods and respective graphical analysis is illustrated in figure 8.

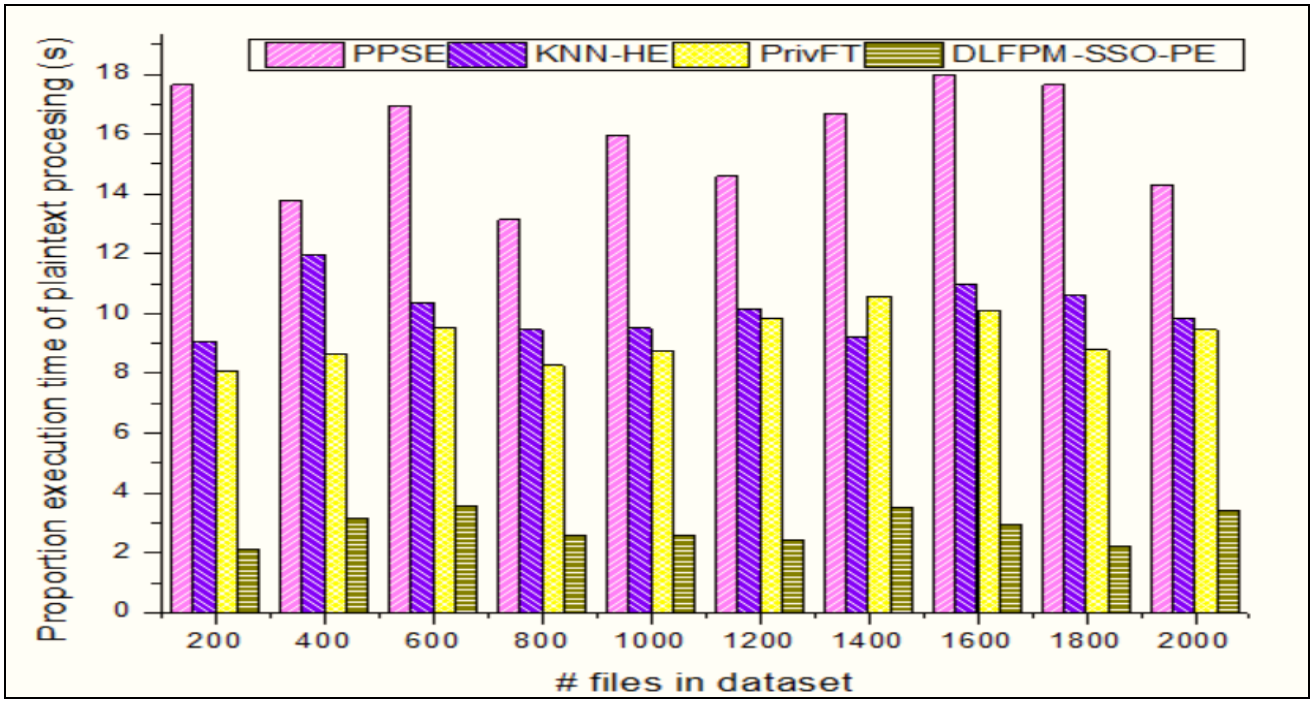


Figure 5: Proportion execution time for plaintext processing

According to the figure 4, the introduced method execute the plaintext with minimum time which means, the method ensure the security very fast compared to existing methods. The plaintext processed by randomly generated integer value r $0 < r < n; and\ r = \mathbb{Z}_{n^2}^*$ and generated keys. For every time, the key values select the number random manner that leads to increase the security because the intermediate users are difficult to guess the key values. More ever, the encryption process uses the homomorphic encryption parameters $cipher\ text\ c = g^m.r^n mod\ n^2$ that increase the plaintext processing time.
Further, the introduced system security efficiency is evaluated in terms of data integrity, confidentiality, authentication and authorization. Then the obtained results are illustrated in table 4.

Table 4: Efficiency Analysis

Methods	Data Integrity	Confidentiality	Authentication	Authorization
PPSE	94.389	93.89	95.39	95.29
KNN-HE	95.28	94.89	95.92	95.92
PrivFT	96.29	96.30	96.20	96.35
DLFPM-SSO-PE	98.9	98.29	98.40	98.20

From the table 4 it clearly shows that introduced DLFPM-SSO-PE method attains maximum security and privacy to the shared intermediate data in third-party server. The obtained efficiency is higher compared to existing methods discussed in review of analysis. According to the results, the obtained graphical result is illustrated in figure 9.

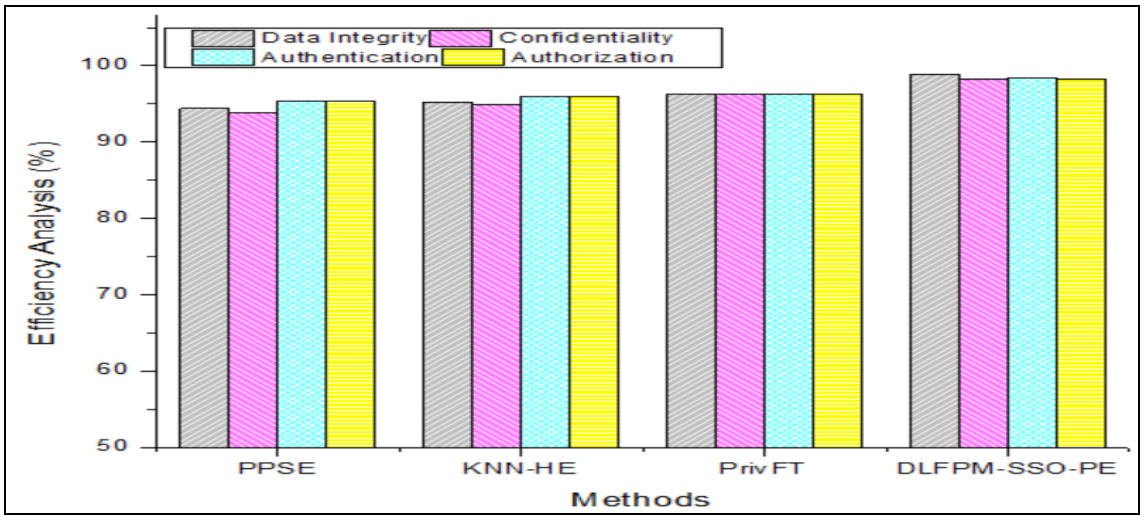


Figure 6 :Efficiency Analysis

Thus the system deep learning network-based frequent pattern mining model and single sign-on associated with Paillier encryption technique (DLFPM-SSO-PE) attains the maximum data security and privacy while sharing data in cloud environment. Around 98.2 to 98.4% of accuracy is ensured in cloud storage process.

5. Conclusion

Thus, the paper analyzing the deep learning network-based frequent pattern mining model and single sign-on associated with Paillier encryption technique (DLFPM-SSO-PE) based secure data storage process in the cloud environment. Initially, the original dataset is analyzed, and an intermediate dataset has to be generated to perform the reuse and research analysis process. The users store their information in the cloud environment by providing the proper data security. Therefore, intermediate data is generated by computing the frequent data patterns. This is achieved by

computing the association rules between data attributes and transactions. Then the extracted patterns are considered as features and processed by long short-term memory deep learning network. The network uses the memory cell for saving and updating every feature involved in this process. This process simplifies the entire sensitive data identification process and resolves the long-term dependency problem. Then the identified information is stored in the cloud storage by providing the SSO user authentication process. Once the user authenticates is performed homomorphic-based private,the public key is generated to encrypt the data. The developed system attains the minimum computation time, cost, and maximum data security accuracy (98.2%) than existing research. In the future, an optimization technique is applied to examine the data to improve the data categorization.

6. References

[1] Bacis, Enrico, Sabrina De Capitani di Vimercati, Sara Foresti, Stefano Paraboschi, Marco Rosa, and Pierangela Samarati. "Access control management for secure cloud storage." In International Conference on Security and Privacy in Communication Systems, pp. 353-372. Springer, Cham, 2016.

[2] Li, Jiaxing, Jigang Wu, and Long Chen. "Block-secure: Blockchain based scheme for secure P2P cloud storage." Information Sciences 465 (2018): 219-231.

[3] Xiong, Shuming, Qiang Ni, Liangmin Wang, and Qian Wang. "SEM-ACSIT: secure and efficient multi-authority access control for IoT cloud storage." IEEE Internet of Things Journal 7, no. 4 (2020): 2914-2927.

[4] Zhang, Xuyun, Chang Liu, Jinjun Chen, and Wanchun Dou. "An upper-bound control approach for cost-effective privacy protection of intermediate dataset storage in cloud." In 2011 IEEE Ninth International Conference on Dependable, Autonomic and Secure Computing, pp. 518-525. IEEE, 2011.

[5] Begum, R. Sabin, and R. Sugumar. "Novel entropy-based approach for cost-effective privacy preservation of intermediate datasets in cloud." Cluster Computing 22, no. 4 (2019): 9581-9588.

[6] Kanagarajan, Suriyalakshmi, and Saranya Vellaichamy. "A heuristic approach for preserving intermediate data set storage on cloud using storage & regeneration cost based on user preference." International Journal of Advances in Engineering & Technology 7, no. 6 (2015): 1829.

[7] Ni, Jianbing, Xiaodong Lin, and Xuemin Sherman Shen. "Toward edge-assisted Internet of Things: From security and efficiency perspectives." IEEE Network 33, no. 2 (2019): 50-57.

[8] Qiu, Meikang, Keke Gai, Hui Zhao, and Meiqin Liu. "Privacy-preserving smart data storage for financial industry in cloud computing." Concurrency and Computation: Practice and Experience 30, no. 5 (2018): e4278.

[9] Yang, Pan, Naixue Xiong, and Jingli Ren. "Data security and privacy protection for cloud storage: A survey." IEEE Access 8 (2020): 131723-131740.

[10] Henze, Martin. "The Quest for Secure and Privacy-preserving Cloud-based Industrial Cooperation." In 2020 IEEE Conference on Communications and Network Security (CNS), pp. 1-5. IEEE, 2020.

[11] Shabir, Muhammad Yasir, Asif Iqbal, Zahid Mahmood, and AtaUllah Ghafoor. "Analysis of classical encryption techniques in cloud computing." Tsinghua Science and Technology 21, no. 1 (2016): 102-113.

[12] Salavi, Rashmi R., Mallikarjun M. Math, and U. P. Kulkarni. "A Survey of Various Cryptographic Techniques: From Traditional Cryptography to Fully Homomorphic Encryption." In Innovations in Computer Science and Engineering, pp. 295-305. Springer, Singapore, 2019.

[13] Li, Ping, Jin Li, Zhengan Huang, Tong Li, Chong-Zhi Gao, Siu-Ming Yiu, and Kai Chen. "Multi-key privacy-preserving deep learning in cloud computing." Future Generation Computer Systems 74 (2017): 76-85.

[14] Kwabena, Owusu-Agyemang, Zhen Qin, Tianming Zhuang, and Zhiguang Qin. "MSCryptoNet: Multi-scheme privacy-preserving deep learning in cloud computing." IEEE Access 7 (2019): 29344-29354.
[15] Li, Xingxin, Youwen Zhu, Jian Wang, and Ji Zhang. "Efficient and secure multi-dimensional geometric range query over encrypted data in cloud." Journal of Parallel and Distributed Computing 131 (2019): 44-54.
[16] Ogiela, Urszula. "Cognitive cryptography for data security in cloud computing." Concurrency and Computation: Practice and Experience 32, no. 18 (2020): e5557.
[17] Idhammad, Mohamed, Karim Afdel, and Mustapha Belouch. "Distributed intrusion detection system for cloud environments based on data mining techniques." Procedia Computer Science 127 (2018): 35-41.
[18] C. Xu, N. Wang, L. Zhu, K. Sharif and C. Zhang, "Achieving Searchable and Privacy-Preserving Data Sharing for Cloud-Assisted E-Healthcare System," in IEEE Internet of Things Journal, vol. 6, no. 5, pp. 8345-8356, Oct. 2019, doi: 10.1109/JIOT.2019.2917186.
[19] S. Sharma, J. Powers and K. Chen, "PrivateGraph: Privacy-Preserving Spectral Analysis of Encrypted Graphs in the Cloud," in IEEE Transactions on Knowledge and Data Engineering, vol. 31, no. 5, pp. 981-995, 1 May 2019, doi: 10.1109/TKDE.2018.2847662.
[20] Yang, J. Xu, J. Weng, J. Zhou and D. S. Wong, "Lightweight and Privacy-Preserving Delegatable Proofs of Storage with Data Dynamics in Cloud Storage," in IEEE Transactions on Cloud Computing, vol. 9, no. 1, pp. 212-225, 1 Jan.-March 2021, doi: 10.1109/TCC.2018.2851256.
[21] Singh, N., Singh, A.K. Data Privacy Protection Mechanisms in Cloud. Data Sci. Eng. 3, 24–39 (2018). https://doi.org/10.1007/s41019-017-0046-0
[22] Dave J., Saharan S., Faruki P., Laxmi V., Gaur M.S. (2017) Secure Random Encryption for Deduplicated Storage. In: Shyamasundar R., Singh V., Vaidya J. (eds) Information Systems Security. ICISS 2017. Lecture Notes in Computer Science, vol 10717. Springer, Cham. https://doi.org/10.1007/978-3-319-72598-7_10
[23] D. N. Wu, Q. Q. Gan and X. M. Wang, "Verifiable Public Key Encryption With Keyword Search Based on Homomorphic Encryption in Multi-User Setting," in IEEE Access, vol. 6, pp. 42445-42453, 2018, doi: 10.1109/ACCESS.2018.2861424.
[24] C. Guo, R. Zhuang, C. Su, C. Z. Liu and K. R. Choo, "Secure and Efficient ${K}$ Nearest Neighbor Query Over Encrypted Uncertain Data in Cloud-IoT Ecosystem," in IEEE Internet of Things Journal, vol. 6, no. 6, pp. 9868-9879, Dec. 2019, doi: 10.1109/JIOT.2019.2932775.
[25] A. A. Badawi, L. Hoang, C. F. Mun, K. Laine and K. M. M. Aung, "PrivFT: Private and Fast Text Classification With Homomorphic Encryption," in IEEE Access, vol. 8, pp. 226544-226556, 2020, doi: 10.1109/ACCESS.2020.3045465.
[26] Mohanty, Sachi Nandan, K. C. Ramya, S. Sheeba Rani, Deepak Gupta, K. Shankar, S. K. Lakshmanaprabu, and Ashish Khanna. "An efficient Lightweight integrated Blockchain (ELIB) model for IoT security and privacy." *Future Generation Computer Systems* 102 (2020): 1027-1037.
[27] Kaaniche, Nesrine, and Maryline Laurent. "Data security and privacy preservation in cloud storage environments based on cryptographic mechanisms." Computer Communications 111 (2017): 120-141.
[28] Shah, Meet, Mohammedhasan Shaikh, Vishwajeet Mishra, and Grinal Tuscano. "Decentralized cloud storage using blockchain." In 2020 4th International Conference on Trends in Electronics and Informatics (ICOEI)(48184), pp. 384-389. IEEE, 2020.
[29] Riad, Khaled, Teng Huang, and Lishan Ke. "A dynamic and hierarchical access control for IoT in multi-authority cloud storage." Journal of Network and Computer Applications 160 (2020): 102633.