INDIAN JOURNAL OF SCIENCE AND TECHNOLOGY

RESEARCH ARTICLE

SABPP: Privacy-Preserving Data Exchange in The Big Data Market Using The Smart Contract Approach

Suman Madan[1]*

1 Jagan Institute of Management Studies, Sector 5, Rohini, Delhi-85, India

OPEN ACCESS

Received: 20-07-2023

Accepted: 03-10-2023

Published: 15-12-2023

Citation: Madan S (2023) SABPP: Privacy-Preserving Data Exchange in The Big Data Market Using The Smart Contract Approach. Indian Journal of Science and Technology 16(46): 4388-4400. https://doi.org/10.17485/IJST/v16i46.1831

* **Corresponding author**.

madan.suman@gmail.com

Funding: None

Competing Interests: None

Published By Indian Society for Education and Environment (iSee)

ISSN
Print: 0974-6846
Electronic: 0974-5645

Abstract

Background: The immense increase of data due to web services, social media, Health care data, and mobile data results in the massive quantity of organized and unorganized data known as big data, which is utilized by various data miners as it contains some sensitive information. **Method:** In this research, a privacy mechanism in the decentralized cloud through the smart contract approach is developed to ensure the privacy of the data and ensure a fair trading strategy. **Findings:** The comparative analysis is revealed in the proposed SABPP model, which shows that the responsiveness attained by the proposed SABPP method is found to be 26.9759sec, 85.2969sec, and 158.6968sec for 20, 60 and 100 nodes respectively. **Novelty:** In this research, the smart contract approach named SABPP is proposed that ensures the smart agreement trading in the Blockchain and overcomes the privacy challenge associated with the trusted third party thereby, ensuring the data availability for the data consumer and privacy for the data provider.

Keywords: Blockchain; smart contract; privacy preservation; authentication; access control; data trading strategy

1 Introduction

Big data is defined as the massive quantity of organized and unorganized data, which is increased by 2.5 Exabytes per day [(1)]. Social media such as Twitter and Facebook, mobile data, YouTube, file hosting websites, digital cameras, GPS signals, healthcare data and some popular web services are responsible for the rapid enhancement in the data volume. Big data is normally classified into three categories such as, Velocity, Variety and Volume. The Velocity is defined as the speed of the data transferred and shared within the network, the Volume is defined as the mass of the data, which exceeds the terabytes and petabytes and Variety is defined as the eruptions of the new information from the mobile computing, machine gadgets and social platforms. Big data Analytics is the process of evaluating the data in social media and other platforms to boost the business organization [(2)]. Hence, it is significant to increase the security and the privacy of the transferred data, as the big data consists of both external data and personal data. The process of preventing delicate information from leakage is known as big data privacy [(3)].

Privacy infringement in domains such as banking, and healthcare leads to catastrophic situations as these domains retain highly confidential person-specified information. The leakage of highly sensitive information to distrustful third parties will create trouble for the users. There are several definitions for data sharing and data privacy, where controlled Information release (CIR) is one among them. It is easy to gather information of the patients in the database by analyzing and consolidating several semi-identifiers such as postcode, age and sex. Hence, just eliminating the aforementioned identifiers from the dataset is not sufficient to preserve the privacy of the individuals. The privacy-preservation is now mandatory in the data mining process to protect the confidential data of the users and the privacy preservation techniques adopted in the data mining domain is known as Privacy-preserving data mining (PPDM) [4,5] . In the current scenario, various researchers make their efforts to restrain the issues related to privacy through advanced technologies such as digital signature [6], Authenticated Data Structure (ADS) [7], message authentication codes [8] and data integrity verification. The ADS is the advanced computing technique, which is utilized to resolve the data authentication issues in the dispersed environment. The ADS verifies the consistency of the data delivered by distrustful third-party services as the validation fails to intervene with the data source. Hence, with the aid of the advanced validation process, the ADS solves the verification issues in the cloud platforms to preserve the privacy of data [9–11].

Machine learning technique received a wide range of consideration as it is widely applicable in artificial intelligence such as financial service, marketing and healthcare to accomplish big data analytic tasks. Data privacy is the most significant and crucial process to be carried out in machine learning, as the General Data Protection Regulation (GDPR) Act prohibits the unauthorized access of the private information of clients as illegal and illegitimate. Hence, the significance of privacy-preserving schemes enhances the requirements of machine learning algorithms. Support Vector Machine (SVM) is considered as the pre-eminent machine learning method widely utilized for the categorization of data. Without being altered by the intense popularity of the deep learning technique, the SVM remains the pre-eminent model as it performs well in both medium-sized data and huge datasets. The highly complicated unpredictable data patterns are evaluated by the SVM with the support of kernel trick [12]. The optimization issues should be restrained in the training stage of the SVM to determine the optimal parameters, whereas in deep learning it has to restrain the non-convex optimization issues. The label for the newly generated data is obtained by estimating the support vector, which comprises the training data and the decision function. The training data and the model parameters are required to be protected to preserve confidentiality [13,14]

This research concentrates on developing a privacy mechanism in the decentralized cloud through smart contract trading to ensure the privacy of the data service provider and the data consumer. In this research, the smart contract trading named SABPP (Smart contract Approach Based Privacy Preservation) model is established to ensure the security of the data in the storage and enable a smart trading experience in the Blockchain environment, which overcomes the privacy challenges associated with the trusted third party. In this smart contract trading, a fair data trading strategy is adopted in the big data market, which ensures payment fairness between the data provider and data consumer, ensuring the data availability for the data consumer, and privacy for the data provider.

The organization of the research article is enumerated as follows: Section 2 explains the motivation and related work analysis along with identified research gaps. The network system model for the research is enumerated in section 3. The proposed privacy-preserving data exchange in the big data market with the smart contract model is elucidated in section 4 and the result analysis is described in section 5. Finally, the conclusion of the paper is illustrated in section 6.

2 Literature Survey

This section motivates the researcher to concentrate in the research and a detailed analysis of the existing methods is deliberated. In [10] presented the optimal geometric transformations for the preservation of privacy in big data known as Privacy preservation of big data via optimal geometric transformations (PABIDOT). Though the PABIDOT excels in execution speed, scalability, attack resistance and accuracy in large-scale privacy-preserving data classification it requires more storage area. In [15] preferred the clustering-based privacy preservation probabilistic model to preserve the privacy of the big data. The clustering method provides better privacy over sanitization-based methods. Yet, locating the most sensitive data-based cluster with additional features is a complex task. In [16] presented the Privacy Preserving Logistic Regression Algorithm (PPLRA) to maintain the privacy of the big data. The main advantage is that the PPLRA can effectively process big data and ensure the privacy of data by utilizing the efficient computing power of the cloud [17] . Yet, the accuracy of the PPLRA scheme is slightly affected by the approximation of the Sigma function. In [18] utilized the perturbation algorithm known as DISTPAB to preserve the privacy of the big data. DISTPAB provides high attack resistance when compared to rotation perturbation and geometric perturbation. Vertical federated learning, which enhances efficiency, is not adopted in the method. In [9] used privacy-preserving adaptive trapdoor hash authentication tree (P-ATHAT) to preserve the confidentiality of the big data. Though the P-ATHAT scheme achieves desirable security, efficiency, and stability it is not applicable for real-time processes as it is time-consuming process. In [19] presented hierarchical fuzzy neural network (PPHFNN) to preserve the confidentiality of the big data. The entire training

procedure is scalable and does not suffer from slow training speeds or gradient vanishing problems. Yet, PPHFNN demands linguistic rules instead of learning examples as prior knowledge.

In [13] presented the SVM model to preserve the privacy of the big data. The SVM-based technique achieved better classification performance for the toy dataset and most of the real datasets. Yet, the system is not suited for high-dimensional data. In [20] presented Blockchain-based privacy preservation, which ensures the credibility of data. Blockchain-based privacy preservation system does not support complex pattern data management and cannot provide general data modeling. In [21] proposed a scheme wherein an information provider consolidates the gathered information and presents the dataset to the huge information market. At that point, the information buyer presents his requests and the market coordinates with the interest of the dataset. Then, the information shopper associates with an information supplier to manage the exchange. To acknowledge the compelling information exchanging huge information market, a few provoke should be settled, including how to guarantee the accessibility of exchanging information, how to secure the protection of personality, and how to guarantee reasonableness. In information exchange, it is difficult for information consumers to confirm the accessibility of the encoded information. In [22] discussed privacy issues instigated because of the usage of blockchain in IoT applications with a focus on the applications of daily use. The strategies for privacy preservation in blockchain-based IoT schemes like encryption, anonymization, private contract, mixing, and differential privacy.

The research gaps highlighted in this research are detailed below:

- Privacy preservation of big data via optimal geometric transformations suffers from the hectic challenge that requires more storage area [10].
- The prime issue experienced in the clustering-based privacy preservation probabilistic model is that the complexity in the Re-construction of modified data [15].
- The accuracy of the PPLRA [16] scheme is slightly affected by the approximation of the Sigma function. Thus, it is considered as the main drawback of the system.
- The prime issue that affects the Hierarchical fuzzy neural network (PPHFNN) [19] is that it demands linguistic rules instead of learning examples as prior knowledge.
- The hectic challenge experienced in the Blockchain-based privacy preservation system does not support complex pattern data management and cannot provide general data modeling [20].
- With regards to the security issue discussed by [21], the supplier is generally hesitant to uncover his genuine character to the information consumer. Besides, the nuclear trade and the installment reasonableness between the information supplier and the information shopper are difficult to ensure.
- Privacy leakage is a hectic issue experienced by blockchain users as it threatens to expose all the transactional details of the user. If any unauthorized person obtains the personal identity of the person, then the confidential information of the user can be leaked. Some of the private information such as transaction amount, which needs to be kept confidential between two end users are leaked due to the availability of the transactional details in the public ledger. [22]

3 System model

The system model of the proposed data exchange in the big data market with smart contract trading is elucidated in this section. The data exchange model in the proposed method is accomplished by four participants such as a) data provider, b) Inter Planetary File System (IPFS), c) Smart contract, and d) data consumers. The system model of the proposed data exchange model is shown in Figure 1.

3.1 Participants in the data exchange method

The brief description of each participant in the proposed data exchange method is demonstrated in the following sub-section.

(1) Data providers: The owner of the data, generally known as data providers generates the list of the data topics to advertise, they register the data in the smart agreement to sell their data.

(2) Smart Contract: The smart contract is the significant platform in the proposed data exchange method, which enables the data transaction between the data providers and the data consumers.

(3) IPFS: The IPFS is the distributed file storage system that possess the address of the data to be exchanged. The IPFS is responsible for the generation of the key in accordance with the address of the data.

(4) Data consumers: The data consumer tends to purchase the data from the data providers by registering to smart contracts through generating the payment transaction.

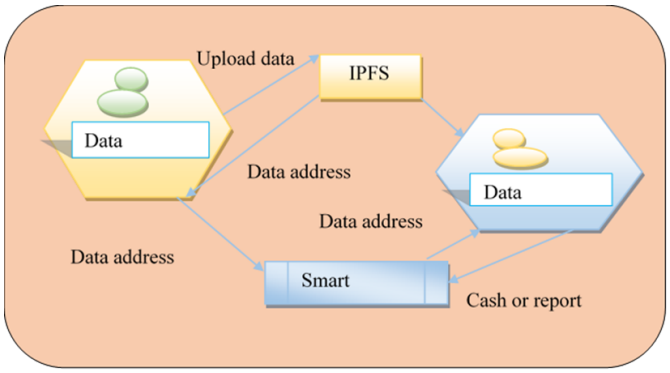

Fig 1. System model of the proposed data exchange in the big data market with the smart contract

3.2 Data exchange model in the big data market

The proposed data exchange model discusses the exchange mechanism between the data provider and consumer, where the data provider vents the original data in the big data market and the data consumer purchases the data. To initialize the trading process, the data provider encrypts the original data and stores it in the IPFS, while acquiring the keys for the stored data. The data storage mechanism promotes the easy transformation of the data by splitting the large-sized data into simple keys. Hence, in the proposed data exchange model, the data consumers upload the data into the IPFS, which provides the address and the key to the data providers. The data provider encrypts the obtained keys and then, uploads the encrypted key in the smart contract. Now, the data consumer purchases the data by acquiring the key from the smart contract by initiating their payment and downloading the data from IPFS. Thus, the data exchange between the data provider and the consumer is performed smoothly through the smart contract without the need for the third party. Moreover, the data consumer utilizes the key to validate the purchased data and upon some issues in the data, the issues are reported to the smart contract. However, there are some possibilities for the potential threats in the data exchange model, which is categorized into three classes, unauthorized data provider threats, fraudulent consumer threats, and intruder attacks.

(1) Fraudulent data provider threats: These threats emerge due to the fraudulent activities of the data provider. For instance, if the product provided by the data provider is found to be entirely different from the advertised data topic.

(2) Fraudulent data consumer threats: These threats are emerged due to the fraudulent activities of the data consumer such as denial of payment transactions.

(3) Intruder attacks: Sometimes the intruder impersonates themselves as the provider or consumer to steal the data from the big data market.

Hence, to avoid these fraudulent activities and fraudulent third-party intrusion, the smart contract is established in this research. Both the provider and consumers need to deposit a specific amount to the smart contract, which is repayable at the end of the successful transaction to avoid fraudulent activities between them.

4 Proposed privacy-preserving data exchange in the big data market with the smart contract

The prime intention of the research is to develop the privacy-preserving data exchange model in the big data market to ensure secure communication between the provider and consumer in the big data market. The smart contract model is established in the research to achieve autonomy, fairness and to avoid the intrusion of third parties in the data exchange. The implications in the smart contracts, which are accepted by both the providers and the consumers, are executed without any alterations and intrusion. The oblivious transaction protocol characteristic is the significant essence of the proposed data exchange model, which enhances the confidentiality and the fairness of the data transfer. The oblivious transaction protocol ensures the privacy of the private and confidential data to such extent and it avoids the fraudulent activities of the provider and consumers. The building block of the proposed data exchange model is briefly elaborated in the following subsection.

4.1 Ether Drafts

In this research, Ether drafts are employed instead of the token system for smooth data exchange of the data between the provider and consumer. The aspects of the Ether Drafts is primarily reflected in the two angles: First, the consumer doesn't have to pay tokens for each inception of an exchange, and the consumers just need to move the tokens from the reserving smart agreement to exchanging smart agreement through the Ether check. Secondly, when the exchange terminates as a result of some strange activities, the tokens straightforwardly reserve the smart agreement, and the consumers utilize the discounted tokens in the reserving smart agreement for the succeeding data exchange. Furthermore, the significant transaction information of consumer and provider are documented in the ether Drafts, which helps them to avoid fraudulent activities. The entire transaction process requires two smart agreements, one for reserving the tokens and the other for trading the data. If a data consumer needs to buy the data from providers, the consumer needs to transfer adequate tokens to the smart agreements to reserve the tokens before the data exchange. The consumer needs to generate the random number $\mathfrak{R}$as, $H_{\mathfrak{R}}$ with the public key of providers as, $E_{K_a^{pub}}(\mathfrak{R})$. Then, the ether draft generated by consumer is mathematically expressed as,

$$D = E\left[K_a^{pub} \parallel A \parallel d \parallel E_{K_a^{pub}}(\mathfrak{R}) \parallel H_{\mathfrak{R}}\right] \oplus \chi \tag{1}$$

where K_a^{pub} is the public key of the data provider, A is the number of tokens paid to the provider, d is the date in which the draft is generated, $E_{K_a^{pub}}(\mathfrak{R})$ is the encrypted random number generated by the consumer, χ refers to the encrypted key of the consumer, and H_R is the hash function of the random number. The consumer needs to sign the draft as,

$$D_{ba} = sig\left(D, K_b^s, \chi\right) \tag{2}$$

The $sig()$ represents the digital signature algorithm and K_b^s represents the private key of the consumer. Thus, the provider attains the draft from the contract agreement to initiate the data trade, and the provider utilizes the public key of the consumer to validate the signature of the draft. The provider needs to decrypt the $E_{K_a^{pub}}(R)$ using the private key K_a^s and obtain the random number $\mathfrak{R}'$. The provider submits $\mathfrak{R}'$to the smart agreement, in which the hashed function of the random number $H_{\mathfrak{R}}$, is verified. The new hashed function is compared with previous hash function in order to determine whether the draft belongs to provider. After the successful validation process, the smart agreement sends the tokens to the providers' address.

4.2 Transaction setup

This section briefly elucidates the specific transaction process of the proposed data exchange model for both the data providers and consumers.

4.2.1 Transaction process

To initiate the process, the data providers upload the data along with security deposit S_a to the smart contract. To establish the *db* fair trade, the consumer needs to pay the security deposit S_b to smart contract. In the first phase, the data provider obtains the data blocks *db* from the database *D* by categorizing the data into *N* number of blocks. In the second phase, data provider uploads the data blocks to the IPFS to attain data address along with data key. In the third phase, the data provider sends the security deposit S_a and block price to the smart contract. If the security deposit is lower than, the agreed amount then, the smart contract rejects the request of the provider. In the fourth phase, the data provider transfers its public key to the smart contracts. In the fifth phase, prior to transaction, the data consumers need to store the adequate tokens on the smart contract for reserving and prepare the payment draft. In the sixth phase, the data consumer transfers the security deposit $S_a$to the data contract. In the seventh phase, the data consumer transfers the block number, which represents the numbers of block, the data consumer needs to purchase. In the eight phase Data consumers transfers its public key to the smart contract. In the ninth phase the data providers utilize public key of the consumer to encrypt the data key. In the final phase, the data provider generates l symmetric keys K^{sym} to establish the smart contract. The smart contract rejects the request of the data providers and the consumers until the contract obtains the determined security deposit.

4.2.2 Initialization of the proposed protocol implementation and transaction

In the proposed data exchange model, the data are dissected into several independent fragments and hybrid encryption standard is adopted in the proposed data exchange model for ensuring the security for the data. The hybrid encryption standard is performed using the Elliptic curve-diffi-huffman (ECDH) algorithm that ensures the security for the data in the smart contract scheme named proposed SABPP protocol.

(a) Establishment phase: The smart contract executes the setup phase to generate the public specifications and the responsible key. The establishment stage is controlled by data provider and let us consider the input as λ, which is the security parameter that establishes the master key m in addition with the public factor P. As the public factor is open-access, the data provider distributes the factor P throughout the databases and blockchain contract systems. The master key m is encrypted by the data provider and the master key m is stored into the transactional blockchain. As mentioned before, the Ether cheques are the smart contracts in the transactional blockchain to such an extent that the smart contract records the encoded keywords and provides a successful search administration for the data consumer. The public factor P and the master key m generated by the data provider is formulated as,

$$P = H(\lambda||r) \otimes \beta \tag{3}$$

$$m = \lambda \otimes a \tag{4}$$

The security factor and the variable r is concatenated, which is hashed and EX-ORed with the random security factor β, where r is the random number ranging between $[0, 1]$. Finally, the master key is generated through EX-ORing the security factors λ and a, where a is the security variable that ranges between 0 and 1. The encrypted master key, m_{en} is given as,

$$m_{en} = \varepsilon(m||a)\, mod\left(\frac{N}{n}\right) \tag{5}$$

Finally, the master key is safeguarded by encrypting the master key using the ECDH algorithm. The security factor a and master key are concatenated, and the concatenated measure is secured through the ECDH-based encryption, which is further multiplied with the module operation on the random number n and data blocks N. The provider stores the master key to the transactional blockchain database, where the EX-OR operation of the database D_1 and encrypted master key is performed. In this stage, the data consumer requests the provider with the password and ID, which is recorded by the provider as, Pp^* and I^* that is re-forwarded to the smart contract to store the user's name and password. The provider initiates the session password as Pp_{sav} for the consumer in order to ensure the authenticity. The consumer receives the session password and records the password as $\overset{P^*_{sav}}{*}$ and reverts back to data provider after fulfilling the identity. The provider then verifies the Pp_{sav} and relegates a quality set λ to the consumer. Additionally, the value-based record address of the consumer is approved and saved in the smart contract. The establishment phase is shown in Figure 2.

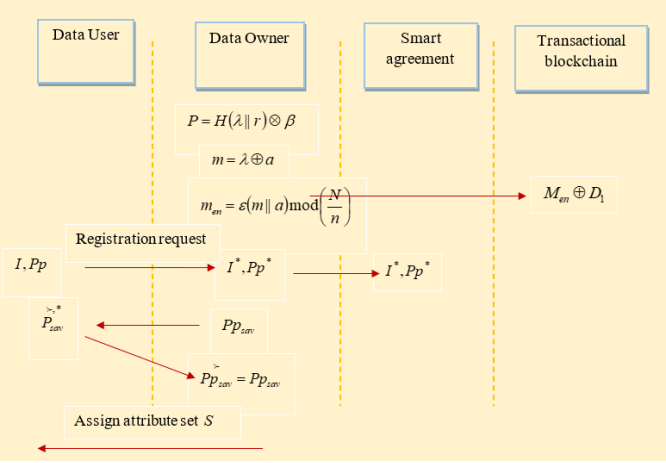

Fig 2. Establishment phase in proposed SABPP model

(b) Registration phase: Both the data providers and the consumers need to be registered with the smart contract to initiate the data transaction. The data provider selects the random $\Re$ within the range of 0 to l and calls the generate function (gen.fn) to establish the key pairs. The provider then calls the sign. KGen to provide the security to the public and private keys. The data provider then summits key pairs and the security deposit to the smart contract. Meanwhile, the consumer produces encrypted

pair keys and submits the private key to the smart contracts. The detail explanation is given in Figure 3. The registration phase is executed by the provider using the secret key k with the attribute set and the master key m. Accordingly, the security parameter is formulated as,

$$z = m \oplus H\left(B||n||\lambda\right) \tag{6}$$

$$k = 8z^4 - 8z^2 + 1 \tag{7}$$

The attribute set is then combined with the arbitrary number and the combined output pertains with hashing function. The variable z is formed by performing the Ex-or operation with combined output. The Chebyshex polynomial is utilized in the variable z to generate the secret key k and the secret key is saved as k^*.

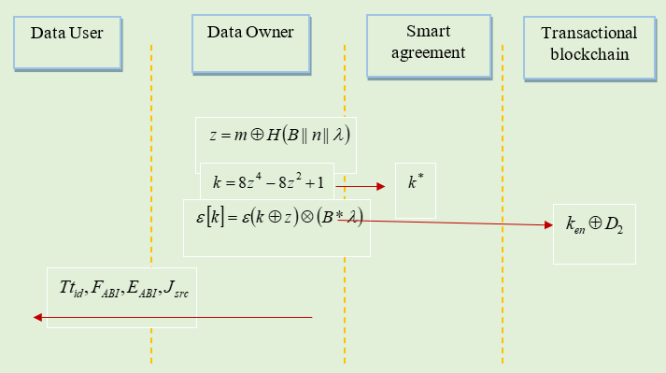

Fig 3. Registration phase of SABPP model

The encoded secret key is formulated as,

$$\varepsilon\left[k\right] = \varepsilon\left(k \oplus z\right) \otimes \left(B * \lambda\right) \tag{8}$$

The EX-OR function is executed with the Chebyshey variables and secret key, which is encrypted and EX-ORed with the security experience factor B. The conditional blockchain is encrypted using the secret key $\varepsilon[k]$ and stored in the smart agreement. The provider then forwards the transaction identity Tt_{id}, address of the contract F_{ad}, Contract Application Binary Interface E_{ABI}, and source code of the contract J_{src} to the consumer through the secure channel.

(c) Encryption phase: The encryption stage is executed by the data provider, and the encryption stage holds three phases, like data encode, key encode, and keyword index. At first, the encryption of the data is performed using the keyword set and the data provider selects the keyword set from information database D and chooses the key utilizing ECDH algorithm to encode the information. Besides, the encrypted data is given as,

$$ED = \varepsilon\left(D||S_k\right) \oplus K_k \tag{9}$$

The data folder is linked with the key word set, which is encrypted and the EX-OR function is actuated with the folder encrypted data and encryption key. The data providers transfer the encoded information to IPFS, which gets the encrypted data and send back the location of the data folder D_l to the data providers. At the subsequent stage, the code text for the metadata is developed by data providers utilizing ED_l and EP. Furthermore, the encrypted data location is determined as,

$$ED_l = \varepsilon\left(D_l||K_k\right) \oplus a \tag{10}$$

The location of the data folder is linked with the folder encryption key and afterward, they are encoded with the aid of encryption function. In any case, the encrypted data is subjected to EX-OR function with the security factor a. The encrypted key is then generated by the data providers and it is represented as

$$EP = \varepsilon\left(P \oplus a\right)||K_k \tag{11}$$

The Ex-OR function is accomplished to α and the public variable and the EX-OR function is then encrypted with the help of folder encryption key. The data provider now generates the code text metadata and it is represented as,

$$C_m = \varepsilon\left(ED_l||EP\right) \oplus S_r \tag{12}$$

Both the encrypted key and the location of the encrypted data are concatenated to form the concatenated data, in which the encryption function is pertained. The ECDH algorithm is utilized to select the random key, which provides the code text metadata while Ex-or with encrypted data. The data provider transfers the code text to the block chain, which obtains the code message and installs the code text with the data set. After the validation of data exchange, the conditional block records both the random key and transaction. The last phase of the encryption stage is the keyword formation. The data provider creates the encrypted keyword folder with respect to the keyword set and a, which is represented as

$$KI_{en} = r||\varepsilon\left(S_k||a\right) \tag{13}$$

The keyword set and the variable a are connected and pertain to the encryption function. The keyword index is generated by the concatenation of the factor r and the encrypted data.

(d) Publish phase: The data provider generates the list of the topics L_T and randomly selects the public key in the given list L and computes the hash function of list and the public key. The data provider generates the signature sig_a and embeds it on topic transaction.(e) Querying phase: The data consumers explore the IPFS to sort out their topic of interest in the data market. Once the consumer sort out their relative topic, they enquire about the content of the topic in the list L_T. The data provider selects the AES private key to encrypt the message and forward the message to the data consumer.

(f) Signature phase: If the inquiry algorithm returns 1 then, the data provider assigns the Sign function to generate the first signature sig_1. The signature is encrypted with the public key of the consumer and sends forward to the consumer. The authentication of the consumer is established by comparing the key generated by the consumers and the key saved in the smart agreement.

(g) Validation phase: The user file is validated by consumer in the validation phase through secret key, random number and the data user ID, which is represented as

$$w = Y \oplus H\left(I \otimes (n||k)\right) \tag{14}$$

A factor is generated by concatenation the secret key and the random number and then, interpolated with the ID of the data users. The generated factor is then exposed to the hashing function and the success factor provided by the smart agreement and the hashed function are now authorized to perform EX-OR function. The validation factor is produced by consumer and advances it to the smart agreement for approving the data consumer. The validation factor is obtained by the smart agreement and saves it as,

$$w^* = Y \oplus H\left(I^* \otimes (n||k^*)\right) \tag{15}$$

The smart agreements confirm w with w^*, on the off-chance that it coordinates, and the client is approved by the smart agreement.

(h) Payment phase: After the validation of the signature, the data consumer initiates the conditional payment and sets the time limit to the payment.

(i) Ensuring Signature: Now, the data provider generates the second signature Sig_2 and calls the smart agreement to execute the transaction. Once it calls the transaction process then, the consumer utilizes both the signature and ensuring signature to extract the private key in order to decrypt the contents.

(j) Decryption phase: The decryption stage is controlled by data consumer, considering both the document encryption key with the recovered data folder and folder encryption key to decrypt the information to create terminal data folder. The data consumer transfers w and Y to data providers, which obtain the document and recorded in the data providers, which is expressed as,

$$\widetilde{w} = Y^* \oplus H\left(I^* \otimes (n||k)\right) \tag{16}$$

The secret key and random number are connected and the output factor is inserted with the data client ID, which are pertained to the hashing capacity. At last, the Ex-or function is executed with Y^*. The data providers confirm whether the validated document produced by data user is coordinated with the document recorded in data provider. Data provider transfers the encrypted key and location of the encrypted data to the data user. The data user saves the location of encrypted data and encoded key and

transfers them to IPFS in addition with A^{S_k}. The IPFS obtains the data transferred from the data user and saves it in IPFS. The encrypted data is now transferred to data user by IPFS. The file transferred by IPFS is downloaded by the data user and decrypt the recovered information utilizing the folder encryption key, which is communicated as,

$$D = de\left(D_R \oplus K_k\right) \tag{17}$$

The folder encryption key is Ex-ORed with recovered data and the resultant variable is decrypted with the help of decryption function.

$$D = D_R||K_k \tag{18}$$

The data file is obtained by concatenating the folder encryption key and recovered data.

$$D = D_R \tag{19}$$

Finally, data required by the data consumer is recovered from the storage unit upon proper authentication by the smart agreement.

(k) Report: The consumer reports with the smart agreement if the keys do not encrypt all the contents in the purchased data. The smart agreement utilizes the responsible key K^{res} to find out the identity of the provider and transfers the security deposit of the provider to the consumer.

4.3 Security requirements

The accompanying security prerequisites ought to be accomplished in a blockchain-based information exchanging protocol.

a) Confidentiality: An un-authorized member, for example, an information customer without the deposit can't acquire the information.

b) Anonymity: To save privacy, an information exchanging protocol should ensure the secrecy of the information supplier. An information purchaser executes information exchanging doesn't have the foggiest idea about the genuine personality ID of the information supplier.

c) Availability: Assuming the information supplier acts truly, the information purchaser can get the information. If the data provider is found to be involved in fraudulent activities such as providing invalid data, then the 'he' cannot obtain his deposit back.

d) Fairness: The fairness in the data exchange is considered as fair transaction in which the provider receives his or her payment, while the consumer receives the valid data.

e) Accountability: The smart agreement will disclose the identity of the provider, who delivers the invalid data to the consumer.

5 Results and Discussion

This section elucidates the evaluation of the privacy-preserving data exchange in the big data market with the smart agreement. The experiment is implemented in the PYTHON and the system framework of the implementation comprised of PYTHON software running in Windows 10 Operating system with 8GB RAM Internal memory

5.1 Comparative method:

The conventional methods utilized for the comparative analysis of the proposed model includes DISTPAB [18] , P-ATHAT [9] , and PPLRA [10].

5.2 Performance matrices

a) Responsiveness: Responsiveness is defined as the ability of the system to execute the allocated task within the specific time

b) Genuine user detection: The genuine user detection is defined as the ability of the system to identify the genuine users from the total number of users in the nodes.

c) Illegitimate user: Illegitimate user detection is defined as the ability of the system to identify the unauthorized users in the network.

5.3 Comparative analysis

This section elucidates the comparative evaluation of proposed privacy-preserving data exchange in the big data market and the competitive methods.

5.3.1 Comparative evaluation using 20 nodes in the network

Figure 4 enumerates the comparative analysis of competent methods in terms of Responsiveness, Memory usage, Genuine Detection and illegitimate user. From the Figure 4 it illustrates that at 20 numbers of users the responsiveness achieved by the proposed SABPP and the competent methods a such as DISTPAB, P-ATHAT and PPLRA are 25.7809 sec, 26.3003 sec, 28.1565 sec and 28.2170 sec, which is increased to 26.9759sec, 28.3241sec, 29.4041sec and 30.3575sec respectively at 100 nodes. The Genuine user detection observed in proposed SABPP and the competent methods such as DISTPAB, P-ATHAT and PPLRA are 95.0000%, 77.9655%, 48.1552% and 45.8621% at 20 numbers of nodes. The legitimate user detection percentage using the proposed SABPP and the competent methods such as DISTPAB, P-ATHAT and PPLRA are 5.0000%, 22.0345%, 51.8448% and 54.1379% when a total of 20 users are communicating in the network.

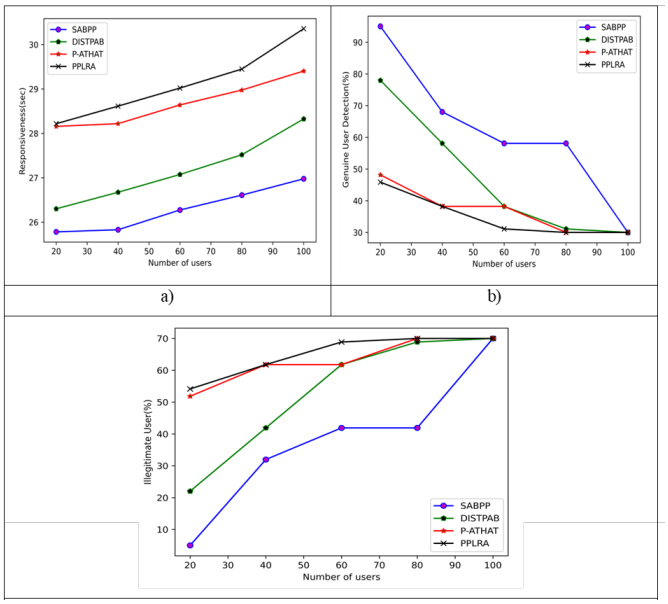

Fig 4. Comparative analysis using 20 nodes of network a) Responsiveness, b) Genuine User detection, c) Illegitimate Users

The comparative analysis of competent methods in terms of Responsiveness, Memory usage, Genuine Detection and illegitimate user is elucidated in the Figure 5. From the Figure 5 it illustrates that at 20 numbers of users the responsiveness achieved by the proposed SABPP and the competent methods a such as DISTPAB, P-ATHAT and PPLRA are 78.5807sec, 80.5768, 81.0674sec and 86.3122sec, which is increased to 85.2969sec, 88.0051sec, 88.4807sec and 94.6012sec respectively at 100 nodes. The Genuine user detection observed in proposed SABPP and the competent methods such as DISTPAB, P-ATHAT and PPLRA are 56.5139%, 56.5139%, 42.7477% and 34.8217% at 40 numbers of users. The legitimate user observed in the proposed SABPP and the competent methods such as DISTPAB, P-ATHAT and PPLRA are at 40 numbers of users are 43.4861%, 43.4861%, 57.2523% and 65.1783%.

Figure 6 depicts the comparative analysis of competent methods in terms of Responsiveness, Memory usage, Genuine Detection and illegitimate user. From the Figure 6, the responsiveness achieved by the proposed SABPP and the competent methods, such as DISTPAB, P-ATHAT and PPLRA are 158.6968sec, 161.7315sec, 162.0727sec and 173.2400sec with 20 nodes, and the responsiveness is further increased to 9750.33sec, 9936.78sec, 9957.75sec and 10643.87sec, respectively for the aforementioned methods when there are 100 nodes simulated in the network. The Genuine user detection observed in proposed SABPP and the competent methods such as DISTPAB, P-ATHAT and PPLRA are 46.2776%, 44.7059%, 42.5045% and 42.0670% at 30 numbers of nodes. The legitimate user observed in the proposed SABPP and the competent methods such as DISTPAB, P-ATHAT and PPLRA are at 30 numbers of users are 53.7224%, 55.2941%, 57.4955% and 57.9330%.

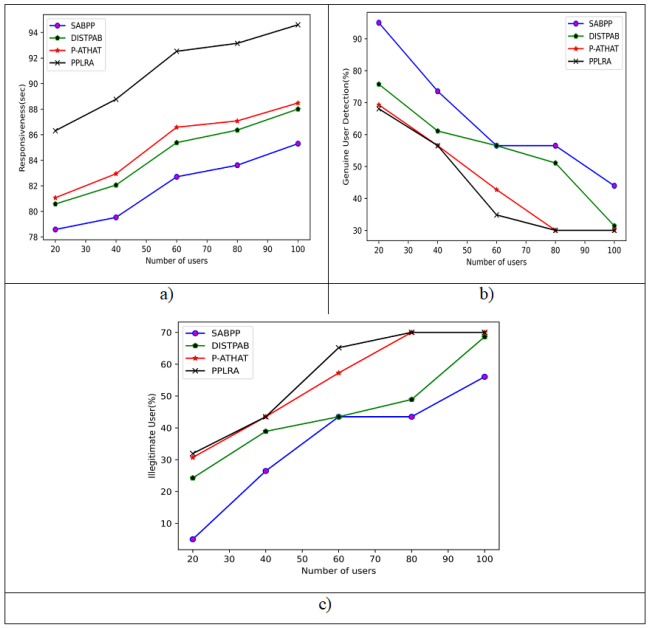

Fig 5. Comparative analysis for using 60 nodes a) Responsiveness, b) Genuine User detection, c) Illegitimate Users

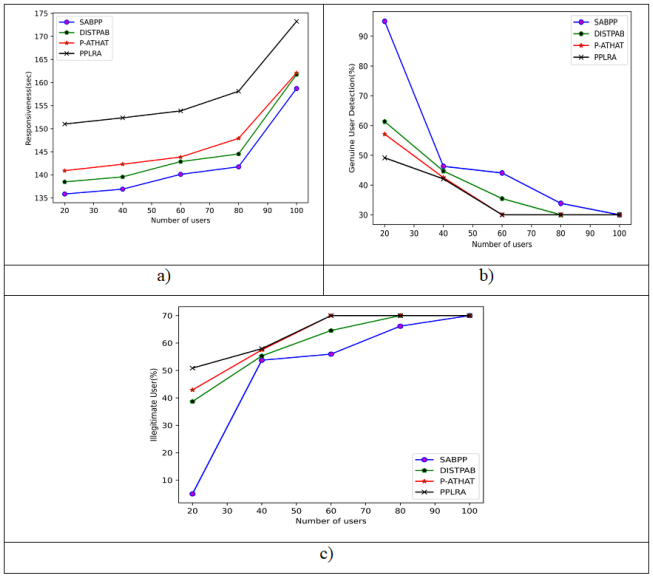

Fig 6. Comparative analysis using 100 nodes a) Responsiveness, b) Genuine User detection, c) Illegitimate Users

Table 1. Comparative discussion of comparative method

Methods	Responsiveness (sec)			Genuine user detection (%)			Illegitimate user (%)		
	20	60	100	20	60	100	20	60	100
Proposed SABPP	**26.9759**	**85.2969**	**158.6968**	**95.0000**	**95.0000**	**95.0000**	**41.9080**	**43.4861**	**53.7224**
DISTPAP	28.3241	88.0051	161.7315	77.9655	75.7569	61.3154	61.7816	43.4861	55.2941
P-ATHAT	29.4041	88.4807	162.0727	48.1552	69.2968	57.1048	61.7816	57.2523	57.4955
PPLRA	30.3575	94.6012	173.2400	45.8621	68.0597	49.1436	68.8793	65.1783	57.9330

5.3.2 Comparative discussion

This section illustrates the comparative discussion of the comparative methods. Table 1 illustrates the peak values attained by proposed SABPP and the competent methods a such as DISTPAB, P-ATHAT and PPLRA. The responsiveness attained by the proposed SABPP method is found to be 26.9759sec, 85.2969sec and 158.6968sec for 20, 60 and 100 nodes respectively. The genuine user detection obtained by the proposed SABPP for 20, 60 and 100 nodes are 95.0000%, 95.0000% and 95.0000% respectively. The illegitimate user achieved by the proposed SABPP for 20, 60 and 100 nodes are 41.9080%, 43.486%, 53.7224% respectively

6 Conclusion

In this research, a privacy mechanism in the decentralized cloud through the smart contract approach is proposed in order to ensure the privacy for the data service provider and the data consumer. In this research, the smart contract approach is utilized to overcome the privacy challenge associated with the trusted third party. In this smart contract approach, a fair data trading strategy is adopted in the big data market, which ensures the payment fairness between the data provider and data consumer, ensuring the data availability for the data consumer, and privacy for the data provider. The experimental analysis is carried out so as to reveal the efficiency of the proposed technique. The responsiveness attained by the proposed SABPP method is found to be 26.9759sec, 85.2969sec and 158.6968sec for 20, 60 and 100 nodes respectively. The genuine user detection obtained by the proposed SABPP for 20, 60 and 100 nodes are 95.0000%, 95.0000% and 95.0000% respectively. The illegitimate user achieved by the proposed SABPP for 20, 60 and 100 nodes are 41.9080%, 43.486%, 53.7224% respectively. The experimental analysis demonstrates that the proposed method exceeds all the other conventional method.

References

1) Khan S, Iqbal K, Faizullah S, Fahad M, Ali J, Ahmed W. Clustering based Privacy Preserving of Big Data using Fuzzification and Anonymization Operation. *International Journal of Advanced Computer Science and Applications*. 2019;10(12):282–289. Available from: https://thesai.org/Publications/ViewPaper?Volume=10&Issue=12&Code=IJACSA&SerialNo=39.
2) Madan S, Bhardwaj K, Gupta S. Critical Analysis of Big Data Privacy Preservation Techniques and Challenges. In: Madan S, Bhardwaj K, Gupta S, editors. International Conference on Innovative Computing and Communications;vol. 1394 of Advances in Intelligent Systems and Computing. Springer, Singapore. 2021;p. 267–278. Available from: https://doi.org/10.1007/978-981-16-3071-2_23.
3) Deebak BD, Al-Turjman F. Privacy-preserving in smart contracts using blockchain and artificial intelligence for cyber risk measurements. *Journal of Information Security and Applications*. 2021;58:102749. Available from: https://doi.org/10.1016/j.jisa.2021.102749.
4) Chamikara MAP, Bertok P, Khalil I, Liu D, Camtepe S. Privacy preserving distributed machine learning with federated learning. *Computer Communications*. 2021;171:112–125. Available from: https://doi.org/10.1016/j.comcom.2021.02.014.
5) Madan S, Goswami P. k-DDD Measure and MapReduce Based Anonymity Model for Secured Privacy-Preserving Big Data Publishing. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*. 2019;27(02):177–199. Available from: https://doi.org/10.1142/S0218488519500089.
6) Yadav V, Agrawal H, Kazi A, Shah Y. Improving Security Through Hashing. *IOSR Journal of Computer Engineering (IOSR-JCE)*. 2020;22(2):47–51. Available from: https://www.iosrjournals.org/iosr-jce/papers/Vol22-issue2/Series-2/I2202024751.pdf.
7) Soleymani SA, Anisi MH, Abdullah AH, Ngadi MA, Goudarzi SH, Khan MK, et al. An authentication and plausibility model for big data analytic under LOS and NLOS conditions in 5G-VANET. *Science China Information Sciences*. 2020;63(12). Available from: https://doi.org/10.1007/s11432-019-2835-4.
8) Yang Z, Wang W, Huang Y. Ensuring reliable logging for data accountability in untrusted cloud storage. In: 2017 IEEE International Conference on Communications (ICC), 21-25 May 2017, Paris, France. IEEE. 2017. Available from: https://ieeexplore.ieee.org/document/7997109.
9) Sun Y, Liu Q, Chen X, Du X. An Adaptive Authenticated Data Structure With Privacy-Preserving for Big Data Stream in Cloud. *IEEE Transactions on Information Forensics and Security*. 2020;15:3295–3310. Available from: https://ieeexplore.ieee.org/document/9063421.
10) Chamikara MAP, Bertok P, Liu D, Camtepe S, Khalil I. Efficient privacy preservation of big data for accurate data mining. *Information Sciences*. 2020;527:420–443. Available from: https://doi.org/10.1016/j.ins.2019.05.053.
11) Chaudhury S, Dhabliya D, Madan S, Chakrabarti S. Blockchain Technology: A Global Provider of Digital Technology and Services. In: Building Secure Business Models Through Blockchain Technology: Tactics, Methods, Limitations, and Performance. IGI Global. 2023;p. 168–193. Available from: https://doi.org/10.4018/978-1-6684-7808-0.ch010.
12) Gaye B, Zhang D, Wulamu A. Improvement of Support Vector Machine Algorithm in Big Data Background. *Mathematical Problems in Engineering*. 2021;2021:1–9. Available from: https://doi.org/10.1155/2021/5594899.
13) Park S, Byun J, Lee J, Cheon JH, Lee J. HE-Friendly Algorithm for Privacy-Preserving SVM Training. *IEEE Access*. 2020;8:57414–57425. Available from: https://ieeexplore.ieee.org/document/9040596.
14) Madan S, Goswami P. Hybrid privacy preservation model for big data publishing on cloud. *International Journal of Advanced Intelligence Paradigms*. 2021;20(3-4):343–355. Available from: https://doi.org/10.1504/IJAIP.2021.119022.
15) Khan S, Iqbal K, Faizullah S, Fahad M, Ali J, Ahmed W. Clustering based Privacy Preserving of Big Data using Fuzzification and Anonymization Operation. *International Journal of Advanced Computer Science and Applications*. 2019;10(12):282–289. Available from: https://doi.org/10.14569/IJACSA.2019.0101239.
16) Fan Y, Bai J, Lei X, Zhang Y, Zhang B, Li KC, et al. Privacy preserving based logistic regression on big data. *Journal of Network and Computer Applications*. 2020;171:102769. Available from: https://doi.org/10.1016/j.jnca.2020.102769.
17) Madan S, Goswami P. Adaptive Privacy Preservation Approach for Big Data Publishing in Cloud using k-anonymization. *Recent Advances in Computer Science and Communications*. 2021;14(8):2678–2688. Available from: https://doi.org/10.2174/2666255813999200630114256.

18) Chamikara MAP, Bertok P, Khalil I, Liu D, Camtepe S. Privacy preserving distributed machine learning with federated learning. *Computer Communications*. 2021;171:112–125. Available from: https://doi.org/10.1016/j.comcom.2021.02.014.
19) Zhang L, Shi Y, Chang YC, Lin CT. Hierarchical Fuzzy Neural Networks With Privacy Preservation for Heterogeneous Big Data. *IEEE Transactions on Fuzzy Systems*. 2021;29(1):46–58. Available from: https://ieeexplore.ieee.org/document/9186813.
20) Guo L, Xie H, Li Y. Data encryption based blockchain and privacy preserving mechanisms towards big data. *Journal of Visual Communication and Image Representation*. 2020;70:102741. Available from: https://doi.org/10.1016/j.jvcir.2019.102741.
21) Li T, Ren W, Xiang Y, Zheng X, Zhu T, Choo KKR, et al. FAPS: A fair, autonomous and privacy-preserving scheme for big data exchange based on oblivious transfer, Ether cheque and smart contracts. *Information Sciences*. 2021;544:469–484. Available from: https://doi.org/10.1016/j.ins.2020.08.116.
22) Hassan MU, Rehmani MH, Chena J. Privacy preservation in blockchain based IoT systems: Integration issues, prospects, challenges, and future research directions. *Future Generation Computer Systems*. 2019;97:512–529. Available from: https://doi.org/10.1016/j.future.2019.02.060.