International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10
Article Received: *16 August 2023* ***Revised:*** *08 October 2023* ***Accepted:*** *21 October 2023*

KBD-Share: Key Aggregation, Blockchain, and Differential Privacy based Secured Data Sharing for Multi-User Cloud Computing

Badugu Ranjith Kumar[1], Dr. M. Victor Jose[2]
[1]Research Scholar, Department of CSE, Noorul Islam Centre for Higher Education, Kanyakumari, TamilNadu, India
ranjithbadugu@gmail.com
[2]Professor/CSE, Vel Tech Multi Tech, Dr.Rangarajan Dr. Sakunthala Engineering College(Autonomous), *Avadi, Chennai,* TamilNadu, India
mvictorjose@gmail.com

Abstract - In today's era of widespread cloud computing and data sharing, the demand for secure and privacy-preserving techniques to facilitate multi-user data sharing is rapidly increasing. However, traditional approaches struggle to effectively address the twin objectives of ensuring privacy protection while preserving the utility of shared data. This predicament holds immense significance due to the pivotal role data sharing plays in diverse domains and applications. However, it also brings about significant privacy vulnerabilities. Consequently, innovative approaches are imperative to achieve a harmonious equilibrium between the utility of shared data and the protection of privacy in scenarios involving multiple users. This paper presents KBD-Share, an innovative framework that addresses the intricacies of ensuring data security and privacy in the context of sharing data among multiple users in cloud computing environments. By seamlessly integrating key aggregation, blockchain technology, and differential privacy techniques, KBD-Share offers an efficient and robust solution to protect sensitive data while facilitating seamless sharing and utilization. Extensive experimental evaluations convincingly establish the superiority of KBD-Share in aspects of data privacy preservation and utility, outperforming existing approaches. This approach achieves the highest R^2 value of 0.9969 exhibiting best data utility, essential for multi-user data sharing in diverse cloud computing applications.

Keywords: multi-user, data sharing, key aggregation, blockchaining, differential privacy.

1. INTRODUCTION

The advancements in today's digital landscape have brought about a critical concern regarding the secure sharing of data among multi-user cloud computing [1]. As organizations and individuals increasingly rely on cloud services for storing and processing their data, ensuring the confidentiality, integrity, and privacy of shared information has become paramount [2]. Privacy breaches and data leakage pose significant risks, including identity theft, financial losses, and reputational damage. Confidential business data, personal information, and intellectual property are valuable assets that need to be safeguarded from unauthorized disclosure [3-4]. Moreover, adhering to data protection regulations [5-7].Encryption is vital for safeguarding confidential information from unauthorized access [8]. It ensures that only individuals with the proper encryption key can decrypt and access the data, maintaining its confidentiality. As an effective information protection control, encryption is crucial for handling confidential data.Key Aggregation is a crucial technique that tackles the challenge of efficiently managing and distributing encryption keys among multiple users [9]. It simplifies the complex process of key management, resulting in reduced computational overhead and enhanced scalability.Blockchain [10,36] technology provides a decentralized and tamper-proof ledger for recording and verifying transactions or data exchanges among multiple users. In multi-user cloud computing environments, blockchain ensures data integrity, immutability, and transparency. It enables secure and auditable data sharing, removing the dependency on central authorities and thereby mitigating the risks associated with data manipulation and unauthorized access.Differential Privacy [11] aims to protect individuals' sensitive information while allowing data analysis and sharing. It introduces randomness to the data to prevent the identification of individuals. In multi-user cloud computing environment, differential privacy techniques enable the aggregation and analysis of data from multiple users while preserving individual privacy. It provides a privacy-preserving mechanism that ensures the confidentiality of user-specific information [12]. Several researches have been carried out to combine the above approaches to build powerful frameworks for enhancing multi-user cloud computing [13, 14, 15]. Inspired by these collaborative networks, this research proposes an integral approach for secured data sharing among multiple users combining key aggregation, block chaining and differential privacy called KBD-Share. At the same time, real time security is likely to safeguard the system without causing damage [37].

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10
Article Received: *16 August 2023* ***Revised:*** *08 October 2023* ***Accepted:*** *21 October 2023*

The proposed model is scalable and This paper is structured as follows: Section 2 provides a comprehensive review of related literature, while Section 3 describes the datasets employed in this research. Section 4 presents the proposed KBD-Share approach, and Section 5 presents a detailed analysis of the experimental results. Section 6 concludes the paper with suggestions for future research directions

2. RELATED RESEARCH

A comprehensive analysis of the current research endeavors concerning secure data sharing in multi-user cloud computing is presented in this section. It delves into the specific domains of key aggregation, blockchain, and differential privacy, highlighting their significance in relation to the proposed KBD-Share mechanism. In the realm of cryptography, various mechanisms facilitate key aggregation, each serving distinct purposes and offering unique advantages. Plutus [16] introduced two decades ago is a highly scalable key management approach, empowering users to maintain direct control over file access through clientbased key distribution and customizable security policies. The aggregate key concept enables the consolidation of multiple secret keys into a single potent key, facilitating data sharing and access control [17]. Multi-key homomorphic encryption [18] empowers individual clients to utilize their own encryption keys, enabling the server to perform homomorphic operations on encrypted data. Secure aggregation, grounded in cryptographic schemes encompassing masking, additive homomorphic encryption, and secret sharing, ensures that the aggregated result remains confidential, safeguarding the privacy of participants [19]. Abbas et al. [20] introduced a Blockchain-assisted Secure Data Management Framework (BSDMF) designed for secure sharing of health information within the Internet of Medical Things (IoMT) network. This framework utilizes blockchain technology to ensure data transmission security and effective data management among interconnected nodes, including personal servers, implantable medical devices, and cloud servers. Experimental results on this framework demonstrated the superior performance of the BSDMF, achieving a high accuracy of 97.2% in data sharing compared to other popular approaches.Lucas et al [21] explored the application of blockchain technology that utilizes smart contracts for automating the tracking and verification of energy demand response services, enabling secure and transparent data sharing among stakeholders.

In their research, Huang et al. [22] put forth a blockchain-based framework for vehicular data sharing. This framework utilizes Zero-Knowledge Proof (ZKP) to safeguard the privacy of vehicle identities and ensures data auditability for Trusted Authorities (TAs). In a recent study, Xie et al. [23] introduced the Trusted Execution Environment and Blockchain supported IoT Data sharing System (TEBDS). This innovative system combines on-chain and off-chain methods to meet the rigorous security demands of IoT data sharing networks. Isaja et al. [24] introduced a blockchain-driven Trusted Framework (TF) for secure and efficient sharing of quality-related information in the supply chain business ecosystem. The TF facilitates the adoption of zero-waste value chain strategies by incorporating a well-defined data model called the Process/Product/Data (PPD) Quality Hallmark, an integrated OpenAPI interface, and anidentity management layer. It enables reliable and trusted sharing of quality data among stakeholders within the production chain.Hassan et al. [25] conducted a comprehensive investigation into integrating differential privacy at multiple layers of the blockchain architecture and applying it in specific blockchain scenarios. In a recent case study, Dyda et al. [26] examined the potential of differential privacy in public health surveillance data, aiming to improve information sharing while safeguarding the confidentiality of this data.In their study, Javed et al [27] proposed ShareChain, an architecture designed to securely share medical data while preserving privacy. In their efforts to address privacy protection in distributed data publishing, Gu et al. [28] devised the LDA-DP algorithm for centralized scenarios. This algorithm perturbs within-class mean vectors and scatter matrices with Gaussian noise.

The problem of releasing private data while preserving privacy was addressed by Jälkö et al [29] through the utilization of Probabilistic Modeling (PM). By redefining synthetic data design as model selection, their approach allows for the incorporation of historic information to improve the generated synthetic data. Empirical evidence from an epidemiological study convincingly demonstrates the reliability of reproducing statistical discoveries using the synthetic data. The Conditional Tabular Generative Adversarial Network (CTGAN) [30] has shown promising results in generating synthetic data that effectively addresses the issue of data imbalance during the training of deep learning models. To mitigate privacy risks, it is essential to apply differential privacy approaches in conjunction with CTGAN for quantifying and controlling the privacy guarantees of the generated synthetic data. Qu et al [31] introduced the Generative Adversarial Net enhanced Differential Privacy (GAN-DP) to address the challenge of preserving individual privacy in IoT networks. This approach is tailored to each party's specific requirements. It strikes a balance between data degradation and privacy leakage, improving data privacy in IoT networks.In a related work [32], a hybrid approach employing differential privacy in a GAN model is proposed to protect critical data in Industrial Internet of Things (IIoT) operations [35].

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10
Article Received: *16 August 2023* ***Revised:*** *08 October 2023* ***Accepted:*** *21 October 2023*

3. DATASETS

The evaluation of the proposed KBD-Share framework utilizes two datasets: the IIoT Wind Turbine (IWT) [33] dataset and the Energy Efficiency (EE) [34] dataset. The IWT dataset is obtained from the Microsoft Azure Predictive Maintenance Template and consists of Supervisory Control and Data Acquisition (SCADA) data from wind turbines, comprising 70 attributes and 49,027 observations. The EE dataset contains data from energy analysis of 768 building shapes simulated in Ecotect, with variations in glazing area, orientation and other attributes. The dataset includes 8 features and 768 samples, aiming to predict two real-valued responses. The evaluation of KBD-Share employs a Linear Regression (LR) model, with the LR model trained on specific independent variables. Table 1 provides the details of these variables, which are selected based on empirical evaluation of KBD-Share with different subsets.

Table 1. Description of Datasets

Dataset	Independent Variables	Dependent Variables
IIoT Wind Turbine	WEC average wind speed (X1), WEC average rotation (X2), WEC average Power (X3), WEC average reactive power (X4), WEC averageavailable power from wind (X5), WEC operating hours (X6)	WEC production (Y)
Energy Efficiency	Relative compactness (X1), Surface Area (X2), Wall Area (X3), Roof area (X4), Overall height (X5), Orientation (X6), Glazing area (X7), Glazing area distribution (X8)	Heating load (Y)

4. PROPOSED KBD-SHARE APPROACH

The architecture of the proposed KBD-Share is designed to ensure secure data sharing while leveraging key aggregation, blockchain technology, and differential privacy. This section provides a high-level description of the KBD-Share architecture with interactions among its components and the key stakeholders involved.

This section provides a high-level description of the KBD-Share architecture with interactions among its components and the key stakeholders involved. KBD-Share architecture and the key components and their interactions are described as below and illustrated with Figure 1.

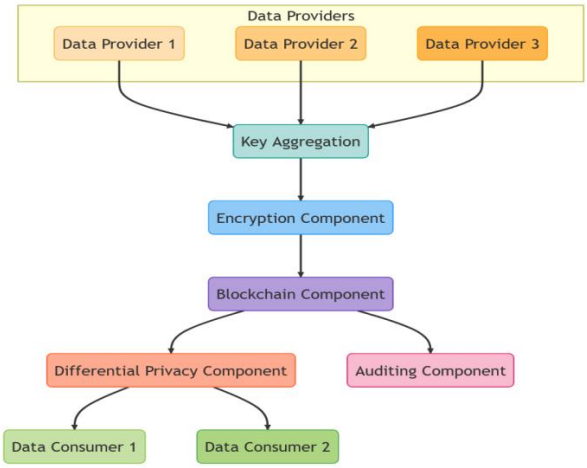

Figure 1. KPD-Share High level Architecture

***a. Data Providers*:**These are entities or users who share their data on the cloud platform. They provide the raw data that needs to be securely shared while preserving privacy.

b. Key Aggregation *Component*: This component is responsible for aggregating keys from multiple data providers to generate a consolidated encryption key. It collects and combines the individual keys securely, ensuring that no single entity has access to the complete key.

c. *Encryption Component*:The Encryption Component utilizes the aggregated key generated by the Key Aggregation Component to encrypt the shared data, employing robust encryption algorithms to guarantee data confidentiality and integrity.

d. *Blockchain Component*: The Blockchain Component acts as a decentralized and tamper-proof ledger, recording access control policies, permissions, and audit logs. It establishes fine-grained access control through the utilization of smart contracts for defining and executing access rules.

e. *Differential Privacy Component:* This component leverages differential privacy techniques to anonymize the shared data while preserving useful statistical properties. It applies privacy-preserving mechanisms to the data to protect individuals' sensitive information.

f. *Data Consumers:* These are authorized users or applications that request access to the shared data. They can query and retrieve the data based on the access control policies defined in the blockchain.

g. Auditing Component: The Auditing Component ensures transparency and accountability in the system. It monitors and records data access events, modifications, and data anonymization processes, providing an audit trail for compliance and security purposes

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10
Article Received: 16 August 2023 *Revised:* 08 October 2023 *Accepted:* 21 October 2023

4.1 Key Stakeholders

The key stakeholders and their roles in the KBD-Share architecture are given as below and illustrated with Figure 2.

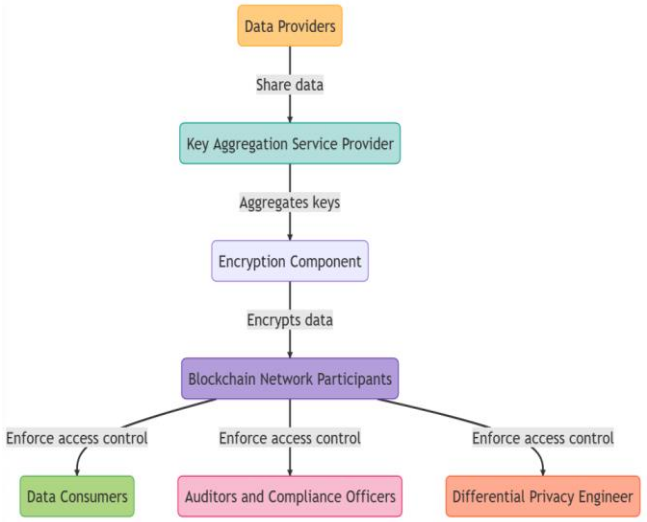

Figure 2. KPD-Share Key Stakeholders

a. ***Data Providers:*** They contribute their data to be securely shared. They play a crucial role in providing access permissions and defining privacy preferences for their data.

b. ***Data Consumers:*** Authorized users or applications that request access to the shared data. They interact with the blockchain component to validate their access rights and retrieve the encrypted data for further analysis or processing.

c. ***Key Aggregation Service Provider:*** The entity responsible for securely aggregating the encryption keys from multiple data providers. They ensure the confidentiality of the individual keys and generate the consolidated key for data encryption.

d. ***Blockchain Network Participants:*** Nodes or entities participating in the blockchain network, responsible for maintaining the decentralized ledger. They validate transactions, enforce access control policies, and contribute to the consensus mechanism.

e. ***Differential Privacy Engineer:*** Experts in differential privacy who design and implement the anonymization techniques used in KBD-Share. They ensure that data sharing adheres to privacy-preserving practices while maintaining data utility.

f.*Auditors and Compliance Officers:* These stakeholders review the auditing logs and ensure that the data sharing process aligns with regulatory and compliance requirements. They verify that access control policies and privacy measures are properly enforced.

4.2 Key Aggregation and Encryption for Secure Data Sharing

This section describes the key aggregation, key generation and distribution, blockchain based access control and the differential privacy-based encryption process used in data sharing.

4.3 key aggregation

The key aggregation technique used in KBD-Share utilizes a secure key aggregation technique to combine individual keys from multiple data providers into a consolidated encryption key.

A set of public keys, denoted as $L = \{pk_1, pk_2, \dots, pk_n\}$, where each public key pk_i corresponds to a data provider i is given as input to the aggregation process. Each public key pk_i is represented as a tuple (y_i, z_i), where $y_i$and $z_i$belong to a cyclic group G. Specifically, y_i is the result of raising the generator$g$ of G to the power of the private key x_i of data provider i, and z_i is the result of raising the element h to the power of x_i. The algorithm outputs the aggregated public key, the consolidated encryption key obtained by aggregating the public keys of all data providers denoted as $apk = (y, z)$, where y and z are elements in the cyclic group G. The key aggregation algorithm is given as below

Algorithm: Key Aggregation

Input:

- $L = \{pk_1, pk_2, \dots, pk_n\}$ -set of public keys of the data providers. For each data provider i where $pk_i = (y_i, z_i) = (g^{x_i}, h^{x_i})$

Output:

- Aggregated Public Key$K_{agg} = (y, z)$

Procedure:

1. Initialization
    - Set y to the identity element e in the cyclic group G
    - Set z to the identity element e in the cyclic group G
2. Key Aggregation
    - For each public key $pk_i = (y_i, z_i)$ in L, do the following:
        - Multiply y by y_i using the group multiplication operation in G
          $y \leftarrow y \cdot y_i$
        - Multiply z by z_i using the group multiplication operation in G
          $z \leftarrow z \cdot z_i$
3. Output
    - Return the aggregated public key $K_{agg} = (y, z)$

4.3.1 Blockchain-based Access Control and Auditing

The utilization of blockchain technology in KBD-Share provides a decentralized and immutable framework for access control. Blockchain serves as a tamper-proof ledger that records access control policies, permissions, and audit logs the data sharing process as below

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10
Article Received: *16 August 2023* ***Revised:*** *08 October 2023* ***Accepted:*** *21 October 2023*

4.3.2 Key Generation and Management

The key management and distribution mechanism in KBD-Share ensures the secure generation, distribution, and management of encryption keys for data sharing, maintaining a high level of security throughout the process. It ensures that the keys are generated securely and distributed only to authorized entities while maintaining their confidentiality and integrity. The key management and distribution mechanism encompasses the following steps.

A. Key Generation:

The necessary cryptographic keys, including public and private keys for each data provider, are generated by a trusted authority, often referred to as the Key Management Center (KMC).

B.Key Distribution:

The public keys are securely distributed to the corresponding data providers using secure channels such as encryption, digital signatures, or secure key exchange protocols. The private keys, which are kept confidential, are securely stored by the respective data providers.

C. Key Updates:

The key management mechanism supports key updates, such as periodic key rotation, revocation of compromised keys, or addition/removal of data providers. Secure procedures and protocols are followed to ensure the secure and efficient update of encryption keys.

D. Key Protection:

The private keys of data providers are protected through measures such as encryption, access controls, and secure storage, safeguarding them against unauthorized access.

E. Key Escrow:

In certain scenarios, key escrow mechanisms may be employed to securely store copies of the private keys. This allows for key recovery or backup options in case of key loss or unavailability of a data provider.

F. Key Revocation:

When a security breach occurs or a data provider's access privileges are revoked, the key management mechanism facilitates the revocation of the corresponding keys to prevent unauthorized access to the shared data.

4.3.3 Blockchain-based Access Control and Auditing

The utilization of blockchain technology in KBD-Share provides a decentralized and immutable framework for access control. Blockchain serves as a tamper-proof ledger that records access control policies, permissions, and audit logs. The access control mechanism in KBD-Share harnesses blockchain technology to assure the integrity and transparency of the data sharing process as below.

Algorithm: Blockchain based Access Control in KBD-Share

Input:

- Blockchain parameters: $B = \{B_1, B_2, \dots, B_n\}$ (Blockchain blocks), $SC = \{SC_1, SC_2, \dots, SC_m\}$ (Smart contracts)
- Data consumer request: Data consumer i

Output:

- Access control decision: $ACD(i)$

Procedure:

1. Validate Blockchain:
   - Perform validation of the blockchain using the validation function $V(B)$. If $V(B)$returns false, terminate the algorithm.
2. Verify Permissions:
   - Retrieve the permissions for data consumer i, denoted as $perm_i$.
3. Access Control Enforcement:
   - For each smart contract $SC_j \in SC$:

Execute the access control enforcement function $E(SC_j)$, which validates and enforces the access control policies defined in SC_j.

If $E(SC_j)$returns false, terminate the algorithm and set $ACD(i)$ as "Access denied."

$ACD(i) \leftarrow$"Access denied."

4. Access Control Decision:
   - If data consumer i satisfies all the access control policies and permissions

$ACD(i) \leftarrow$"Access Granted."

else

$ACD(i) \leftarrow$"Access Granted."

5. Return $ACD(i)$ as the access control decision for data consumer i.

Smart contracts play a vital role in KBD-Share by enforcing access policies. These self-executing contracts have predefined terms written into their code, enabling them to define and enforce the access policies governing the conditions and permissions for accessing shared data as described below.

1. Access Policy Definition:

Each smart contract SC_j defines an access policy using mathematical notations. It specifies the conditions and requirements for granting access to the shared data. The access policy can be represented as P_j, where P_j is a logical expression involving variables and operations.

2. Execution of Smart Contract:

When a data consumer requests access to the shared data, the corresponding smart contract SC_j is executed. The smart

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10
***Article Received:** 16 August 2023 **Revised:** 08 October 2023 **Accepted:** 21 October 2023*

contract evaluates the access policy P_j based on the input parameters and current system state.

3. Access Control Enforcement:

The smart contract enforces the access policy P_j by evaluating whether the conditions specified in P_j are satisfied. This evaluation involves mathematical operations and logical comparisons. Access to the shared data is either granted or completely denied based on whether the access policy is satisfied.

4. Smart Contract Execution Result:

The execution of the smart contract results in a Boolean value, denoted as R_j, which represents the outcome of the access control enforcement. If R_j is true, it indicates that the access policy is satisfied, and access is granted. If R_j is false, it signifies that the access policy is not satisfied, and access is denied.

Auditability and transparency are important aspects enabled by blockchain technology in KBD-Share. The immutability and decentralized nature of the blockchain provide a transparent and tamper-proof record of all access control events and activities. In this research, enforcement of auditability and transparency in KBD-Share is realized as below.

1. Transaction Recording:

Each access control event, modification, and data anonymization process is recorded as a transaction on the blockchain. Let T_i represent the i-th transaction recorded on the blockchain.

2. Access Control Logs:

Access control logs capture details of data consumer requests, permissions, and access control decisions. The access control log can be represented as Log_i, which contains the relevant information for the i-th access control event.

3. Blockchain Validation:

The blockchain can be validated using mathematical functions and algorithms. Let $V(B)$ denote the validation function that verifies the integrity and consistency of the blockchain B. If $V(B)$ returns true, it indicates that the blockchain is valid; otherwise, it is considered invalid or inconsistent.

4. Transparency:

The transparency of the blockchain is achieved through its decentralized nature, allowing all participants to have access to the same set of transaction records. Let TR represent the set of all transactions recorded on the blockchain.

5. Audit Trail:

The blockchain serves as an audit trail, providing a chronological and immutable record of access control events and activities. The audit trail can be represented as AT, which contains a sequence of transactions T_i.

The above process enable the stakeholders to review and verify the access control events and activities recorded on the blockchain, ensuring transparency and accountability in the system.

4.3.4 Leveraging differential privacy for data anonymization

Differential privacy is seamlessly incorporated into KBD-Share to ensure privacy protection during data sharing, anonymizing the shared data and safeguarding individuals' sensitive information. The privacy parameter ϵquantifies the level of privacy protection provided to individuals in the shared data. A lower ϵ value indicates a higher level of privacy assurance.

To achieve differential privacy, noise is added to the query results or data during the sharing process. Let $Q(D)$ represent a query performed on the dataset D. The noise addition process can be defined as in (1), where $\mathcal{N}(\Delta Q, \epsilon)$ denotes the noise added to the query result, ΔQ represents the querysensitivity, and ϵ is the privacy attribute.

$$Q(D) + \mathcal{N}(\Delta Q, \epsilon) \quad (1)$$

The integration of differential privacy in KBD-Share involves privacy and utility trade-offs. As the parameter ϵ decreases to provide stronger privacy guarantees, the utility or accuracy of the shared data may be compromised. Achieving a balance between privacy and utility is a critical consideration in the anonymization process. By integrating differential privacy techniques into KBD-Share, sensitive information in the shared data is protected, and privacy guarantees are provided. When it comes to secure data sharing, the utilization of ε-differential privacy for data anonymization through generalization is instrumental in safeguarding the privacy of the shared data. This approach involves replacing specific values in the dataset with more generalized or coarse-grained representations to mitigate the risk of re-identification. Through the incorporation of ε-differential privacy, a formal guarantee is established to protect the privacy of the shared data, regardless of the inclusion or exclusion of any individual's data, thereby preserving the overall data privacy. The privacy budget ε determines the maximum allowable privacy loss for the shared data. Sensitive attributes, denoted as A, are identified in the dataset, and each attribute $a \in A$ has a generalization hierarchy, $H(a)$, specifying the levels of generalization. The original dataset, D, undergoes the generalization process, where specific attribute values $d(a)$ are replaced with their corresponding generalized values $g(a)$, as defined by the hierarchy $H(a)$. In order to attain ε-differential privacy and bolster the security of the shared data, random noise $N(a)$ is introduced to the generalized values,

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10
Article Received: *16 August 2023* ***Revised:*** *08 October 2023* ***Accepted:*** *21 October 2023*

as outlined in (1). The incorporation of noise adds an additional layer of protection, preventing potential adversaries from inferring sensitive information even when they have partial knowledge of the original data. The anonymized dataset, denoted as D', is then securely shared with authorized users or stored in a trusted environment. During query processing or data analysis, the noise added during anonymization is taken into account to ensure accurate results without compromising individual privacy. Queries on the anonymized dataset are performed using $Q(D)$, considering the perturbed data and preserving the privacy guarantees provided by ε-differential privacy. The privacy guarantees can be analyzed as follows:

a. Privacy Budget: The privacy budget, represented by ε, is the maximum allowable privacy loss that quantifies the information an adversary can learn about an individual by observing the shared data.

b.Neighboring Datasets: Two datasets, D and D', are regarded as neighbors if they vary by the inclusion or exclusion of an individual's data. The analysis considers the datasets D_i and D'_i, which are neighboring datasets differing only in the data of individual i.

c. Privacy Loss: The privacy loss, denoted as L_i, measures the extent to which an adversary can learn about an individual by observing the output of the data sharing process.

d. Differential Privacy Property: The ε-differential privacy property ensures that the probability of observing a specific output from the shared data remains nearly the same, regardless of the presence or absence of an individual's data. This property is expressed as in (2) where, Mrepresents the data sharing process, D is the original dataset, D_i is a neighboring dataset, and O is a specific output of the sharing process. This inequality guarantees that the probability of observing a specific output is minimally affected by the presence or absence of an individual's data, thus preserving privacy.

$$Pr[M(D) = O] \leq \exp(\epsilon) \cdot Pr[M(D_i) = O] \quad (2)$$

e. Privacy Risk: The privacy risk, denoted as δ, quantifies the probability of exceeding the budget ε and serves as an indicator of the likelihood of privacy breaches. A smaller value of δ signifies a lower risk of privacy violations.

f Privacy Loss Bounds: The objective is to ensure that the privacy loss for any individual is bounded by the privacy budget ε. This is expressed as in (3) , indicating that the probability of the privacy loss for individual i exceeding ε is no more than δ.

$$Pr[L_i > \epsilon] \leq \delta \quad (3)$$

The approach ensures robust privacy protection for shared data by carefully selecting generalization hierarchies, incorporating noise mechanisms, and adhering to ε-differential privacy guarantees.

In the data sharing process of KBD-Share, a data block B is securely shared among N users leveraging key aggregation, blockchain, and differential privacy techniques as below Figure 3.

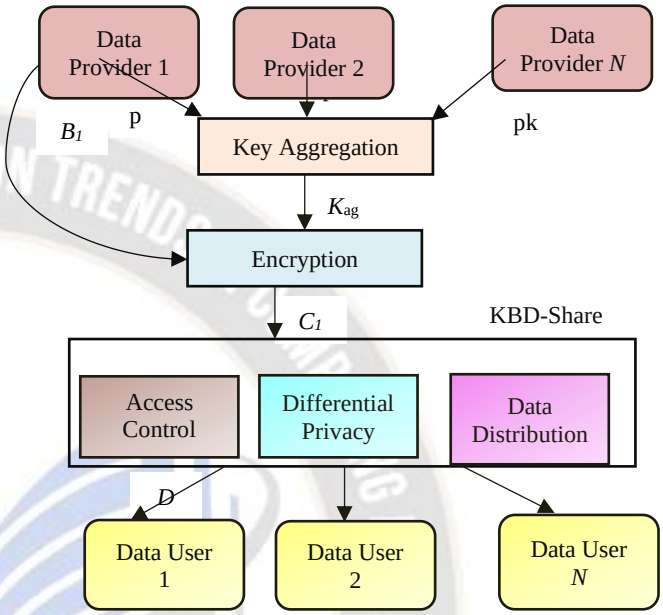

Figure 3. Secured Data Sharing

1. *Key Aggregation:*
   - N data providers contribute their individual encryption keys, denoted as $pk_i \forall i = 1 to N$.
   - The key aggregation component combines these individual keys to generate a consolidated encryption key, denoted as in (4)

$$K_{agg} = \text{Aggregation}(pk_1, pk_2, \ldots, pk_N) \quad (4)$$

2. *Encryption:*
   - For encryption of the data block B, the key K_{agg} is usedas in $C = \text{Encrypt}(B, K_{agg})$, where C represents the encrypted data block.
3. *Blockchain-based Access Control*

The access control is enforced with the following steps.

a. Recording in Blockchain:

The encrypted data block, denoted as C , is saved as a transaction in the blockchain.

The blockchain serves as a distributed ledger, storing blocks that consist of multiple transactions.

For a block B_i, T_i represents the transaction containing the encrypted data block C_i.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10
***Article Received:** 16 August 2023 **Revised:** 08 October 2023 **Accepted:** 21 October 2023*

The process of recording the transaction in the blockchain can be represented as in (5)

$$\text{RecordInBC}(C,\text{AccessControlPolicies},\text{Permissions},\text{AuditLogs}) \rightarrow B_i = \{T_1, T_2, \ldots, T_i\} \quad (5)$$

b. Smart Contracts and Access Control Rules:

Smart contracts are employed to enforce granular access control rules and define the privileges of each user.

Let SC_i represent the smart contract associated with the i-th block in the blockchain.

Access Control Policies (ACPs), permissions, and Audit Logs (ALs) are stored within the smart contract for each transaction.

The access control process can be represented as:

$$\text{AccessControl}(\text{SC}_i,\text{ACPs},\text{Permissions},\text{ALs})$$

c. Access Control Enforcement:

Upon the users 'request for access to the encrypted data block, their access privileges are verified by comparing them to the access control policies specified in the associated smart contract.

Let $User_i$ represent the i-th user requesting access, and A_{user_i} denote the access privileges of $User_i$.

The access control enforcement process can be represented as:

$$\text{ValidateAccess}(\text{SC}_i, User_i, A\text{user}_i)$$

d. Access Grant or Denial:

Based on the validation result, the user is either granted access to the encrypted data block or denied access.

If access is granted, the user can proceed with retrieving the encrypted data block and further decryption.

If access is denied, the user is restricted from accessing the encrypted data block.

This process can be represented as:

$$\text{AccessGranted}(User_i)$$

4. Differential Privacy:

To preserve privacy, differential privacy techniques are applied to the encrypted data block, C.

During the data sharing process, privacy guarantees are ensured by introducing noise to the data or query results.

The privacy mechanism can be represented as $DC_i \leftarrow AddNoise(C, \epsilon_i)$, where DC_i represents the distinct copy of the differentially private encrypted data block for user i and ϵ_i is the privacy parameter specific to user i.

5. Data Distribution:

The differentially private encrypted data block, DC_i, is distributed among the N users through a secure channel or a trusted cloud platform.

Each user receives a copy of the differentially private encrypted data block, but they do not have access to the original data in its unencrypted form.

5.0. EXPERIMENTAL RESULTS AND DISCUSSIONS

This section presents the empirical results on the evaluation of KBD-Share with objective metrics and a comparative analysis of this approach and representative works.

5.1 Experimental Setup

For the experimental evaluation of KBD-Share, the system is configured with an Intel Core i7-8700K CPU, 1 TB SSD and 16GB RAM. The operating system used is Windows 10 Professional, and the implementation is done in MATLAB R2023a. The privacy parameter ϵ is set to 0.1, representing the desired level of differential privacy. The key aggregation threshold is chosen as 4, indicating that at least four user keys are required for key aggregation. Gaussian noise at 0.5 standard deviation and Laplace noise with a scale parameter of 1.0 is added during the data anonymization process. The encryption algorithm employed is AES-256 with CBC mode. MATLAB, along with its Communications Toolbox and Statistics and Machine Learning Toolbox, is used for implementing the secure data sharing process, generating noise, and performing statistical analysis. By adopting this setup, a dependable and consistent platform is established to assess the performance of KBD-Share across privacy guarantees, data utility, and computational efficiency.

5.2 Performance Metrics

The proposed KBD-Share is evaluated with six privacy metrics and one utility metric to quantitatively assess its performance. These metrics provide insights into the effectiveness of the privacy-preserving mechanisms employed in KBD-Share as well as the utility of the differentially private encrypted data for predictive problems. Towards fair evaluation and comparison, KBD-Share is evaluated with the privacy metrics Mean Square Error (MSE), Normalized Variance (VAR), Directed Hausdorff (HAUS), Kullback-Leibler divergence (KL), Procrustes (PRO), and Pearson's Correlation Coefficient (PCC) used in [32], to capture different aspects of the privacy preservation capabilities.

Additionally, the utility metric R-squared score (R^2) is used to assess the extent to which the differential private encrypted data preserves the relationship between the independent and dependent variables. By considering these performance metrics, the evaluation provides a comprehensive analysis of the effectiveness and utility of KBD-Share in terms of privacy preservation and predictive modeling.

In the context of KBD-Share, MSE is utilized as a privacy measure to quantify the difference between the features of the

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10
Article Received: *16 August 2023* ***Revised:*** *08 October 2023* ***Accepted:*** *21 October 2023*

original and test datasets. It is calculated by taking the mean of the squared errors between corresponding feature values as in (6), where x_i and y_i represent the feature value from the original and test data sets, X and Y respectively. Generally, a lower MSE indicates a smaller difference between X and Y, suggesting better privacy preservation.

$$MSE = \frac{1}{n}\sum_{i=1}^{n} (x_i - y_i)^2 \quad (6)$$

VAR is a metric used to compare the dispersion or difference in distribution between X and Y as in (7). A higher NV indicates a greater difference in distribution between X and Y, suggesting a potential loss of privacy. Conversely, a lower NV indicates a smaller difference in distribution, indicating better privacy preservation.

$$VAR = \frac{1}{n}\sum_{i=1}^{n} \left(\frac{x_i - y_i}{\sigma_x}\right)^2 \quad (7)$$

Generally, HAUS is used to measure the dissimilarity or difference between two datasets, X and Y. It quantifies the maximum distance between a point in one dataset and its closest point in the other dataset. In the context of the KBD-Share approach, the HAUS metric is employed to assess the dissimilarity between X and Y as in (8). A higher HAUS value indicates a greater dissimilarity between the datasets, suggesting a stronger level of privacy achieved by the KBD-Share mechanism.

$$\text{HAUS} = \max_{x \in X} \min_{y \in Y} \| x - y \| \quad (8)$$

To measure the information gain or loss between the probability distributions of X and Y, KL divergence is used, which quantifies the dissimilarity between these two distributions. It provides insights into how much information is gained or lost during the transformation from the original dataset to the test dataset, thereby reflecting the level of privacy protection achieved through the differential privacy mechanisms applied in KBD-Share. It is computed as in (9) where $P(i)$ and $Q(i)$ are the probabilities of observing the value i in the original and decrypted datasets, respectively. A larger value of KL divergence signifies a larger dissimilarity, signifying a potential loss of privacy. Conversely, a lower KL divergence value indicates a smaller dissimilarity, indicating better privacy preservation.

$$\text{KL}(P \parallel Q) = \sum_i P(i) \log\left(\frac{P(i)}{Q(i)}\right) \quad (9)$$

Procrustes (PRO) analysis is a statistical method that compares a collection of shapes by transforming them into a state of maximal superimposition. It achieves this by minimizing the sum of squared distances between corresponding points in each shape through affine transformation of their coordinate matrices. Given the datasets dataset X and Y, PRO is computed as in (10), where T is the transformation matrix. The resulting value of PRO indicates the level of similarity between the two sets of shapes. A smaller value indicates a higher degree of alignment and similarity, while a larger value suggests greater dissimilarity.

$$PRO(X, Y) = \min_{T} ||X - YT||^2 \quad (10)$$

The PCC measures the linear correlation between X and Y as in (11) ,where x_i and y_i represent the original and decrypted values, respectively, and $\overline{X}$ and $\overline{Y}$ denote the means of the original and decrypted datasets, respectively. It falls in the range [-1 1], with -1 indicating a perfect negative linear relationship, 0 indicating no linear relationship, and 1 indicating a perfect positive linear relationship between X and Y.

$$PCC(X, Y) = \frac{\sum_{i=1}^{n} (x_i - \overline{X})(y_i - \overline{Y})}{\sqrt{\sum_{i=1}^{n} (x_i - \overline{X})^2 \sum_{i=1}^{n} (y_i - \overline{Y})^2}} \quad (11)$$

R^2 is a utility metric which evaluates how well the model fits the observed data and provides an indication of the percentage of the variability in the dependent variable that can be explained by the independent variables as in (12). In this research, it is used to assess the utility of differential private encrypted data in preserving the relationship between the independent and dependent variables. A higher value of R^2 indicates a stronger alignment of the model with the data, implying that the utility of the data for predictive purposes has not been substantially compromised by the differential privacy mechanisms implemented in KBD-Share.

$$R^2 = \frac{\sum_{i=1}^{n} x_i y_i - (\sum_{i=1}^{n} x_i)(\sum_{i=1}^{n} y_i)}{\sqrt{n(\sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2)}\sqrt{n(\sum_{i=1}^{n} y_i^2 - (\sum_{i=1}^{n} y_i)^2)}} \quad (12)$$

5.3 Privacy Preservation and Prediction Accuracy Evaluation

The performance of the KBD-Share approach is evaluated with the IWT and EE datasets with the above metrics and the results are presented in Table2 and Table 3 respectively for $\epsilon = 0.1$. The two best results are highlighted with bold red and blue faces.

Table 2. Performance Metrics – IIoT Wind Turbine Dataset

No. of Users / Metric	10	20	30	40	50
MSE↓	**0.0432**	**0.0439**	0.0449	0.0446	0.0492
VAR↓	**0.0442**	**0.0443**	0.0444	0.0456	0.0481
HD↓	0.0608	0.0504	0.0590	**0.0364**	**0.0373**
KL↓	0.1042	0.1135	0.1612	**0.1038**	**0.1029**
PRO↓	1.872	**1.8612**	1.8622	**1.847**	1.862
PCC↑	**0.9997**	**0.9996**	0.9995	0.9995	0.9993
R^2↑	**0.9969**	**0.9951**	0.9939	0.9925	0.9901

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10
Article Received: *16 August 2023* ***Revised:*** *08 October 2023* ***Accepted:*** *21 October 2023*

From Table 2, it is observed that the MSE and VAR values consistently remain low across different number of users, indicating effective privacy preservation. Further, these metrics increase with the number of users. While the HD values exhibit fluctuations, indicating diverse shape similarity between the datasets based on the number of users, the KL divergence demonstrates a consistent decrease as the number of users increases, reflecting an enhancement in distribution.

Similarly, Table 3 shows that the MSE values for the EE dataset also remain small, with a significant increase compared to that of the IWT dataset. It is seen that best values of the MSE, VAR, HD, PRO, PCC and R^2 metrics are achieved for 10 users, while degradations are observed with increased number of users. The lowest KL divergence is achieved for 30 users.

Table 3. Performance Metrics – Energy Efficiency Dataset

No. of Users / Metric	10	20	30	40	50
MSE↓	**0.2021**	0.2031	**0.2034**	0.2038	0.2041
VAR↓	**0.2023**	**0.2033**	0.2027	0.2041	0.2037
HD↓	**0.1414**	**0.1435**	0.1473	0.1594	0.1605
KL↓	0.3120	**0.3114**	**0.3108**	0.3124	0.3118
PRO↓	**0.1979**	0.1988	**0.1982**	0.1996	0.1992
PCC↑	**0.8956**	0.8951	**0.8954**	0.8947	0.8949
R^2↑	**0.8021**	0.8012	**0.8018**	0.8004	0.8008

This indicates that the KBD-Share approach effectively preserves privacy while maintaining a reasonable level of utility with both datasets. The results suggest that a smaller number of users tends to yield better performance in terms of privacy preservation, as indicated by lower MSE, VAR, HD, and KL divergence values. Nevertheless, it is crucial to acknowledge that the variations in performance observed across different metrics and datasets underscore the inherent trade-off between privacy and utility. Determining the optimal number of users to strike the rightbalancemay rely on the specific requirements and priorities of the given application.

Table 4. Performance Comparison with State-of-the-Art

Privacy Parameter ϵ	Data Sharing Approach			
	KBD-Share (Proposed)		Hybrid GAN-DP [32] -2023	
	IWT	EE	IWT	EE
0.1	**0.9969**	**0.8021**	**0.9371**	**0.762**
0.2	**0.9958**	**0.8006**	**0.9361**	**0.7605**
0.3	**0.9947**	**0.7991**	**0.935**	**0.7591**
0.4	**0.9936**	**0.7975**	**0.9342**	**0.7577**
0.5	**0.9925**	**0.796**	**0.9331**	**0.7562**

CTGAN [30] -2019		PM [29] -2021		GAN-Enhanced DP [31] (2022)
EE	IWT	EE	IWT	EE
0.7468	0.8368	0.7019	0.7866	0.6528
0.7453	0.8359	0.7006	0.7857	0.6516
0.7439	0.835	0.6993	0.7849	0.6503
0.7425	0.8274	0.6981	0.774	0.6491
0.7411	0.8253	0.6966	0.7732	0.6479

The R^2 values are illustrated with Figure 4 to comprehend the performance of the data sharing approaches based on differential privacy.

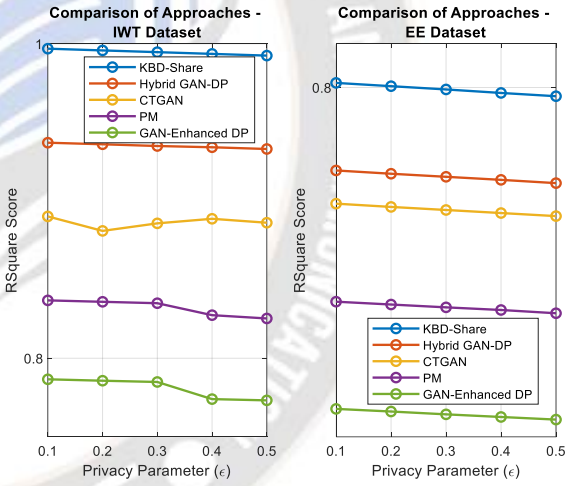

Figure 4. Prediction Accuracy Comparison with the State of the Art

5.4 Discussions

Based on the experimental findings and comparative analysis, it is evident that KBD-Share demonstrates superior performance over other data sharing approaches, surpassing them by a substantial margin. While PM, CTGAN, GAN-DP, and the Hybrid GAN all address the issue of privacy preservation to some extent, KBD-Share provides a more robust and comprehensive solution. When comparing the various approaches for maintaining privacy while releasing private data, it becomes apparent that each method brings distinct contributions to tackle this challenge effectively.

6. CONCLUSION

This paper presents KBD-Share, a comprehensive framework for secure and privacy-preserving data sharing. By incorporating key aggregation, blockchain technology, and differential privacy techniques, KBD-Share provides an effective solution for addressing privacy concerns in data sharing scenarios. The experimental evaluation demonstrated the superiority of KBD-Share in terms of privacy preservation, and utility compared to existing approaches. This approach offers data owners the opportunity to confidently share their data, reducing the likelihood of privacy breaches and striking a favorable balance between privacy and utility. KBD-share can be deployed in various domains and applications where privacy and security of data are paramount.

REFERENCES

[1] Ali, M., Dhamotharan, R., Khan, E., Khan, S. U., Vasilakos, A. V., Li, K., & Zomaya, A. Y. (2015). SeDaSC: secure data sharing in clouds. IEEE Systems Journal, 11(2), 395-404.

[2] Sepehri, M., Trombetta, A., & Sepehri, M. (2017). Secure data sharing in cloud using an efficient inner-product proxy re-encryption scheme. Journal of Cyber Security and Mobility, 339-378.

[3] Federal Trade Commission. (2016). Protecting personal information: A guide for business.

[4] Talesh, S. A. (2018). Data breach, privacy, and cyber insurance: How insurance companies act as "compliance managers" for businesses. Law & Social Inquiry, 43(2), 417-440.

[5] Regulation, G. D. P. (2018). General data protection regulation (GDPR). Intersoft Consulting, Accessed in October, 24(1).

[6] McGraw, D. (2013). Building public trust in uses of Health Insurance Portability and Accountability Act de-identified data. Journal of the American Medical Informatics Association, 20(1), 29-34.

[7] Williams, B. R., & Adamson, J. (2022). PCI Compliance: Understand and implement effective PCI data security standard compliance. CRC Press.

[8] Gupta, I., Gurnani, D., Gupta, N., Singla, C., Thakral, P., & Singh, A. K. (2022). Compendium of data security in cloud storage by applying hybridization of encryption algorithm.

[9] Veeravelli, R. (2018). Data Sharing and Access Using Aggregate Key Concept.

[10] Habib, G., Sharma, S., Ibrahim, S., Ahmad, I., Qureshi, S., & Ishfaq, M. (2022). Blockchain Technology: Benefits, Challenges, Applications, and Integration of Blockchain Technology with Cloud Computing. Future Internet, 14(11), 341.

[11] Hassan, M. U., Rehmani, M. H., & Chen, J. (2019). Differential privacy techniques for cyber physical systems: a survey. IEEE Communications Surveys & Tutorials, 22(1), 746-789.

[12] Tang, P., Chen, R., Su, S., Guo, S., Ju, L., & Liu, G. (2023). Multi-party sequential data publishing under differential privacy. IEEE Transactions on Knowledge and Data Engineering.

[13] Goryczka, S., Xiong, L., &Sunderam, V. (2013, March). Secure multiparty aggregation with differential privacy: A comparative study. In Proceedings of the Joint EDBT/ICDT 2013 Workshops (pp. 155-163).

]14] Pettai, M., & Laud, P. (2015, December). Combining differential privacy and secure multiparty computation. In Proceedings of the 31st Annual Computer Security Applications Conference (pp. 421-430).

[15] Yang, M., Margheri, A., Hu, R., & Sassone, V. (2018). Differentially private data sharing in a cloud federation with blockchain. IEEE Cloud Computing, 5(6), 69-79.

[16] Kallahalla, M., Riedel, E., Swaminathan, R., Wang, Q., & Fu, K. (2003, March). Plutus: Scalable Secure File Sharing on Untrusted Storage. In Fast (Vol. 3, pp. 29-42).

[17] Cui, B., Liu, Z., & Wang, L. (2015). Key-aggregate searchable encryption (KASE) for group data sharing via cloud storage. IEEE Transactions on computers, 65(8), 2374-2385.

[18] Ma, J., Naas, S. A., Sigg, S., & Lyu, X. (2022). Privacy-preserving federated learning based on multi-key homomorphic encryption. International Journal of Intelligent Systems, 37(9), 5880-5901.

[19] Bonawitz, K., Ivanov, V., Kreuter, B., Marcedone, A., McMahan, H. B., Patel, S., ... & Seth, K. (2017, October). Practical secure aggregation for privacy-preserving machine learning. In proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security (pp. 1175-1191).

[20] Abbas, A., Alroobaea, R., Krichen, M., Rubaiee, S., Vimal, S., & Almansour, F. M. (2021). Blockchain-assisted secured data management framework for health information analysis based on Internet of Medical Things. Personal and ubiquitous computing, 1-14.

[21] Lucas, A., Geneiatakis, D., Soupionis, Y., Nai-Fovino, I., & Kotsakis, E. (2021). Blockchain technology applied to energy demand response service tracking and data sharing. Energies, 14(7), 1881.

[22] Huang, J., Kong, L., Wang, J., Chen, G., Gao, J., Huang, G., & Khan, M. K. (2023). Secure Data Sharing over Vehicular Networks Based on Multi-Sharding Blockchain. ACM Transactions on Sensor Networks.

[23] Xie, H., Zheng, J., He, T., Wei, S., & Hu, C. (2023). TEBDS: A Trusted Execution Environment-and-Blockchain-supported IoT data sharing system. Future Generation Computer Systems, 140, 321-330.

[24] Isaja, M., Nguyen, P., Goknil, A., Sen, S., Husom, E. J., Tverdal, S., ... &Lamplmair, P. (2023). A blockchain-based framework for trusted quality data sharing towards zero-defect manufacturing. Computers in Industry, 146, 103853.

[25] Hassan, M. U., Rehmani, M. H., & Chen, J. (2020). Differential privacy in blockchain technology: A futuristic approach. Journal of Parallel and Distributed Computing, 145, 50-74.

[26] Dyda, A., Purcell, M., Curtis, S., Field, E., Pillai, P., Ricardo, K., ... & Lau, C. L. (2021). Differential privacy for public health data: An innovative tool to optimize information sharing while protecting data confidentiality. Patterns, 2(12), 100366.

[27] Javed, L., Anjum, A., Yakubu, B. M., Iqbal, M., Moqurrab, S. A., & Srivastava, G. (2022). ShareChain: Blockchain-enabled model for sharing patient data using federated learning and differential privacy. Expert Systems, e13131.

[28] Gu, Z., Zhang, K., & Zhang, G. (2022). Multiparty Data Publishing via Blockchain and Differential Privacy. Security and Communication Networks, 2022.

[29] Jälkö, J., Lagerspetz, E., Haukka, J., Tarkoma, S., Honkela, A., & Kaski, S. (2021). Privacy-preserving data sharing via probabilistic modeling. Patterns, 2(7), 100271.

[30] Xu, L., Skoularidou, M., Cuesta-Infante, A., & Veeramachaneni, K. (2019). Modeling tabular data using conditional gan. Advances in Neural Information Processing Systems, 32.

[31] Qu, Y., Gao, L., Yu, S., & Xiang, Y. (2022). Personalized Privacy Protection of IoTs Using GAN-Enhanced Differential Privacy. In Privacy Preservation in IoT: Machine Learning Approaches: A Comprehensive Survey and Use Cases (pp. 49-76). Singapore: Springer Nature Singapore.

[32] Hindistan, Y. S., & Yetkin, E. F. (2023). A Hybrid Approach with GAN and DP for Privacy Preservation of IIoT Data. IEEE Access.

[33] W. Soontronchai. (2019). IIoT Data of Wind Turbine. [Online]. Available: https://www.kaggle.com/datasets/wasuratme96/iiot-data-of-wind-turbine [Accessed on 30 May 2023]

[34] A.Xifara. (2012). Energy Efficiency Data Set. [Online]. Available: https://archive.ics.uci.edu/dataset/242/energy+efficiency [Accessed on 30 May 2023]

[35] K. K. Vaigandla, "Communication Technologies and Challenges on 6G Networks for the Internet: Internet of Things (IoT) Based Analysis," 2022 2nd International Conference on Innovative Practices in Technology and Management (ICIPTM), 2022, pp. 27-31, doi: 10.1109/ICIPTM54933.2022.9753990.

[36] R. Yadav, Ritambhara, K. K. Vaigandla, G. S. P. Ghantasala, R. Singh and D. Gangodkar, "The Block Chain Technology to protect Data Access using Intelligent Contracts Mechanism Security Framework for 5G Networks," 2022 5th International Conference on Contemporary Computing and Informatics (IC3I), Uttar Pradesh, India, 2022, pp. 108-112, doi: 10.1109/IC3I56241.2022.10072740.

]37] Badugu Ranjith Kumar,M.. Victor Jose (2023). A Survey on Various Secure Data Sharing Methods in Multi-User Cloud Using Key Aggregation and Blockchain Based Approaches A Frame Work,International Conference on Networking and Computer Applications (ICNA 2023), Noorul Islam Centre for higher Education, Kanyakumari, Tamil Nadu,Conference Proceedings ISBN No. 978-81-945891-3-6.