

## RESEARCH ARTICLE

# Optimizing Industrial IoT Data Security Through Blockchain-Enabled Incentive-Driven Game Theoretic Approach for Data Sharing

MUHAMMAD NOMAN SOHAIL<sup>1</sup>, ADEEL ANJUM<sup>2</sup>, IFTIKHAR AHMED SAEED<sup>1,3</sup>,  
MADIHA HAIDER SYED<sup>2</sup>, AXEL JANTSCH<sup>4</sup>, (Senior Member, IEEE),  
AND SEMEEN REHMAN<sup>4</sup>, (Member, IEEE)

<sup>1</sup>Department of Computer Science, The University of Lahore, Sargodha Campus, Lahore 40100, Pakistan

<sup>2</sup>Institute of Information Technology, Quaid-i-Azam University, Islamabad 45320, Pakistan

<sup>3</sup>Koblenz University of Applied Sciences, 56075 Koblenz, Germany

<sup>4</sup>Institute of Computer Technology, Technische Universität Wien (TU Wien), 1040 Vienna, Austria

Corresponding author: Axel Jantsch (axel.jantsch@tuwien.ac.at)

**ABSTRACT** Connecting smart industrial components to computer networks revolutionizes business operations. However, in the Industrial Internet of Things (IIoT), the sharing of data has bandwidth, computational, and privacy issues. Researchers presented cloud computing and fine-grained access control to overcome these challenges. However, traditional centralized computing systems involve single points of failure. To mitigate these challenges, we have proposed a secure and incentive-based data-sharing framework for IIoT systems using blockchain technology. We leverage blockchain due to its ability to provide secure and tamper-resistant data storage and sharing as participants store their data on a distributed ledger (DL), preventing unauthorized access. A security protocol is designed that utilizes the properties of elliptic curve cryptography (ECC). Moreover, Shapley value is employed to calculate revenue and distribute it fairly. To perform the formal security evaluation, we have conducted extensive simulations using the Automated Validation of Internet Security Protocols and Applications (AVISPA) and Scyther protocol simulation tools, which demonstrated that our protocol is robust against various adversarial attacks. The experimental results show that the proposed incentive distribution framework demonstrated fairness in the distribution of revenue among participants.

**INDEX TERMS** Data sharing, game theory, profit distribution, elliptic curve, industrial IoT.

## I. INTRODUCTION

The IIoT is a complicated system made up of interconnected smart industrial components and computer platforms [1]. The goal of IIoT is to monitor industrial processes to improve overall system performance. In an industrial setting, a significant number of devices, such as sensors and actuators, generate a large volume of data [2] that is used for better decision-making and to maintain productivity and improve efficiency within the industry. Thus, the heterogeneous nature of data from various sources is frequently offloaded to

the cloud for analysis purposes. To share data with users, the traditional framework faces some challenges, such as bandwidth or computational overheads leading to inaccurate analysis, resulting in poor decision-making and economic losses as well as the efficiency of the system [3]. Moreover, there may be a threat to data in terms of privacy preservation due to unauthorized access. Different methods have been used to mitigate privacy and security threats [4], in which either the sender has to use security protocols to secure the data transmission or leaves no choice for the receiver to trust the data [5], [6]. Though some existing schemes resolve some issues, there is still a threat posed by participants in the sharing system regarding trusted and accurate data sharing [7].

The associate editor coordinating the review of this manuscript and approving it for publication was Stefano Scanzio<sup>1</sup>.

To address privacy concerns, many researchers presented various solutions comprised of different techniques, but many emphasized the use of a cloud computing environment to create platforms to share data in a secure way on customer demand [8], [9]. Various security models, such as deploying fine-grained access control mechanisms, have been proposed to ensure secure data storage in a cloud environment [10].

Cloud computing is a centralized environment, and all the data resides at a single point for storage and processing. Traditional centralized data-sharing systems have a single point of failure and raise grave privacy concerns. Hence, these challenges require a secure and trustworthy platform that provides data security and privacy protection and prevents attackers from acquiring and disclosing information about the data or system participants. Recent research trends [11], [12], [13] show that blockchains provide reliable security to data storage and sharing that significantly reduces privacy threats during data sharing by controlling unauthorized access to data via access control measures [14]. Overall, blockchain integration in IIoT enhances data security and provides identity verification for the participants [15]. A fair incentive distribution mechanism is important for a collaborative environment that prevents manipulation of data sharing for individual gain. Similarly, Maintaining data privacy in collaborative settings requires controlling access rights to secure individual data.

Our objective is to evaluate how blockchain technology might enhance IIoT systems, focusing on data sharing. The decentralized and tamper-proof nature of blockchain technology helps to achieve trust and consensus for resource sharing among the participants. The security features of blockchain such as precisely recording the transactions on the distributed ledger (DL) create a trusted environment that meets the need for a data-sharing framework. A consortium blockchain helps to control data access using an access control mechanism. It manages user's identities to limit unauthorized access to data so that shared data is only visible to authorized participants of the system. So, to increase data availability by preserving privacy blockchain provides an efficient solution.

In addition, to guarantee the delivery of accurate and trustworthy data an incentive distribution mechanism is needed that fairly distributes incentives among data providers. For this purpose, in our research, we show that Shapley value estimates the contribution of each participant of the consortium and achieves fairness during incentive distribution based on provided data. Thus, we proposed a consortium blockchain-based incentive distribution framework for secure data sharing that uses the game theoretic approach Shapley value to solve the consortium incentive distribution problems.

Our secure framework integrates blockchain and Shapley value to design a collaborative environment to exchange data between different participants of the system. Our study addresses two main issues in industrial data-sharing systems. First, an attacker may obtain and misuse data. A trustworthy

platform that guarantees privacy and prevents attackers from accessing and modifying data is essential. Equal profit distribution among data suppliers is a second concern when sharing data with consumers [16]. This research provides a safe, incentive-based IIoT distribution mechanism that uses blockchain to store data providers' data on a secure DL that limits unauthorized access. We also employed HMAC and ECDH hashing methods to prevent unauthorized transmission of data between parties.

We have summarized our contributions as follows:

- We propose an efficient and anonymous authentication framework for data-sharing system participants that utilizes HMAC and ECC for anonymity and integrity. It resists impersonation, replay, and secret disclosure attacks. In addition, it secures participant data transmission and protects from unauthorized access.
- We used permission blockchain to develop an efficient ecosystem that provides a resources-sharing platform to ensure consensus and trust among the participants. It guarantees the visibility of data to the authorized participants and protects the privacy of sensitive data.
- We have developed an incentive distribution framework based on the Shapley value [17] for data sharing among multiple participants. We consider the collaboration of participants and develop the revenue distribution model.
- We have conducted extensive simulations to verify the performance of the proposed incentive distribution framework against the factors that affect the distribution of revenue among the participants.

The remaining paper is as follows: **II** presents a summary of existing work, **III** describes the preliminaries, **IV** describes problem formulation, in **V** and **VI** proposed methodology is presented, **VII** and **VIII** described the performance evaluation and conclusion of the paper, respectively.

## II. REVIEW OF RELATED LITERATURE

In this section, we have briefly described the existing work in two sections and summarized in Table 1.

### A. SECURITY PROTOCOLS FOR SECURE DATA SHARING

There are various solutions proposed to address the security challenges by using blockchain technology, papers [18] propose an authentication protocol for cross-domain IoT device interaction using Merkle tree structure to store sensitive information. Similarly in [19] author proposes an authentication protocol for secure cross-domain data exchange in IIoT. However, this paper does not provide a potential solution to incorporate the large number of IIoT devices. To solve the scalability and security problems in cross-domain networks various blockchain-based frameworks are proposed in [20], [21], [22], and [23]. In [20] Wang et al. proposed a scheme that addresses efficiency and security challenges in cross-domain IIoT by using edge servers to assist smart devices in achieving cross-domain authentication, while the lightweight message

authentication algorithm guarantees message security with low computational overhead. In [24] the author presented a lightweight authentication protocol based on ECC for fault-tolerant wireless sensor networks. It provides secure communication between resource-constrained devices for data sharing. Though it provides the minimum trade-off between communication and computation complexities, it does not consider the load balancing and adversarial attack scenario. In [25] author proposes a security framework for IIoT that leverages ECC, hashing, bitwise XOR operation, and PUF to protect the system against physical attacks and address the challenges of resource-constrained IIoT environment. Saleem et al. [26] used hashing, the PUF condition, and bitwise operations to construct a vehicular network data security protocol. Although computationally inadequate, it protects the confidentiality of personal data. Wang et al. proposed an IIoT privacy solution in [27]. This framework uses blockchain technology to record transactions securely and safeguard sensitive data during transmission. In [28], the author proposed a key aggregate searchable encryption (KASE) data-sharing technique for fog-enabled IoT environments. It preserves data confidentiality, integrity, and availability while allowing authorized users to communicate data in a secure setting [29], [30]. Yi et al. [31] developed an authentication mechanism for wireless sensor networks (WSNs) in IIoT to secure data. However, IIoT systems need scalability, usability, and computational overhead. The authors created a blockchain-based data-sharing approach in [32] and [33] to solve security and privacy issues with decentralized data. Zero-knowledge proofs isolate data providers from their shared data to avoid data tampering. Tanveer et al. introduced a lightweight and efficient authentication protocol for IIoT in [34] to solve security problems such as limited resources and rapid authentication processes [35].

The motivation behind this research is to enhance the security of data-sharing systems in IIoT which can be implemented in two ways such as actual and simulation-based implementation. To secure the underlying system three different types of cryptographic methods can be used: symmetric encryption, asymmetric encryption, and hybrid encryption which uses both symmetric and asymmetric encryption techniques. Among these techniques, ECC provides promising security solutions for resource-constrained devices due to its characteristics of efficient use of bandwidth and computation. These properties make ECC suitable for IIoT devices with minimal processing power and communication.

### B. INCENTIVE DISTRIBUTION MECHANISMS

Zhang et al. [36] presented a smart contract-based quality-driven incentive system for secure data exchange across IoT devices. Blockchain verifies data integrity, while smart contracts apply the incentive system. Mai et al. [37] introduced a federated learning auction technique. Data owners offer their data and processing resources to users via

a double-auction method. The proposed approach balances data owners and consumers while maintaining dependability and effectiveness. Chen et al. [38] developed COMSA to tackle profit distribution concerns in micro-edge computing and ensure users get high-quality end-to-end service. In [39] and [40], the author's strategy considers spectrum allocation and data routing for best service quality, but it does not address possible security and privacy issues during the double auction [37], [41], [42]. Kang et al. [43] address the challenges of privacy and data correctness in the domain of the healthcare metaverse. The author aims to provide an efficient solution to counter the privacy issue using a decentralized model. In addition, it promotes active user participation and collaboration by incentivizing the process that encourages the data owners to provide truthful data.

Our main contribution to this research domain is to develop a fair incentive distribution mechanism using Shapley value that fairly distributes the profit among the data owner. There are some solutions already been proposed using Shapley value for different industries. Li and Qu [44] and Yang et al. [45] provide a solution to distribute profit fairly to improve supply chain management. Similarly, in other domains, research incorporated Shapley value to improve the performance of the system. Dang et al. improve the accuracy and collaboration using Shapley value by incentivizing cooperation among the clients [46]. Chai and Zeng [47] proposed a Shapley value-based computation offloading framework in edge computing.

## III. PRELIMINARIES

### A. ELLIPTIC CURVE

Suppose a large prime number  $p > 3$  that defines the finite field  $F_q$  and  $4a^3 + 27b^2 \neq 0$  only if a group of points  $a, b \in F_p$ ,  $EC_p$  is the elliptic curve that satisfies these points on curve  $EC_p(a, b) : y^2 = x^3 + ax + b \text{ mod } P$ . Our technique is based on the one-way Elliptic Curve Discrete Logarithm Problem (ECDLP), which is hard to compute. Understanding the cyclic group and its features helps create Discrete Logarithm Problems (DLPs).  $G$  is cyclic if  $G \Leftrightarrow \exists \alpha \in G$  and  $\text{ord}(\alpha) = |G|$ .  $\alpha$  has the same cardinality as the basic element group  $G$ . The DLP requires cyclic groups with closure, associativity, identity, and inverse. DLP in a cyclic group  $G$  to find  $x$  such as  $\alpha^x \equiv \beta \text{ mod } p$ .

### B. ECDLP

In a cryptosystem, the finite groups or cyclic groups play an essential role in building the structure that is considered during the construction of any system. To define DLPs more precisely, we first need to understand the cyclic group and its properties. A  $G$  group is cyclic,  $G \Leftrightarrow \exists \alpha \in G$ , such that  $\text{ord}(\alpha) = |G|$ . It means that  $\alpha$  has the same cardinality as the group  $G$ . A group must satisfy the following properties in order to use it for constructing the DLPs.

- *Closure*: A group is closed,  $\forall a, b \in G$  such that  $aob = c$ .

TABLE 1. Summary of existing work.

Ref	Problem Statement	Proposed Solution	Cryptographic and Incentive Technique	Limitation
[18]	Privacy protection in cross-domain data sharing system	Blockchain-based authentication management system	Utilized Merkle tree structure to store sensitive information	Requires high computational cost
[19]	Security and privacy protection during data transmission in IIoT	Authentication protocol to ensure message authenticity and privacy protection	Utilized hashing, Chebyshev polynomial operations and bitwise-XOR	Does not support untraceability and unlinkability
[20]	Efficiency and security challenges in IIoT	Lightweight message authentication framework for edge nodes in IIoT	Utilized ECC and hash function	Does not provide robust protection against advanced attacks
[21]	Critical security vulnerabilities in centralized system	Hyperledger based framework to enhance security	Utilized NuCypher threshold re-encryption mechanism	It has high communication overhead
[22]	To address the challenge of sensitive data privacy in IIoT	Permissioned blockchain based framework to protect the privacy	Bilinear Pairing and computational Diffie-Hellman Problem	Does not support scalability and mutual authentication
[23]	Security challenges in resource-constrained devices in IIoT	Symmetric key based authentication framework for key agreement	Utilized hash functions and XOR operations	Does not provide security against traceability and unlinkability attacks
[24]	Communication and computational complexities for secure data sharing	Lightweight authentication framework	Utilized ECDH and ECC	Does not provide scalability and adaptability
[25]	Eliminate the security threats to provide anonymity and untraceability	Provable secure authentication framework in IIoT	Utilized PUF, bitwise XOR, and ECC	Does not support interoperability
[26]	Focus on reducing computational overhead and privacy protection	Physically secure key agreement protocol in Ad-networks	PUF, hash functions, and fuzzy extractor	Lack scalability feature, does not support unlinkability
[27]	To securely share the private information in smart factories	Blockchain-based security solution for privacy protection in IIoT	Intelligent Elliptic Curve Digital Signature (IECDSA)	Does not support forward secrecy and length extension attack
[28]	To protect the data sharing process from adversarial attacks	Blockchain-based key-aggregation scheme for data communication	Utilized bilinear pairing	Require high computational cost, does not support unlinkability
[29]	Protect communication of sensitive information against security threats	Three-factor based secure and anonymous authentication protocol for WSN	Utilized PUF and hashing	Does not support unlinkability and traceability
[30]	Ensures the data privacy from exposure during cloud storage	Efficient authentication protocol for IIoT based WSN	Utilized PUF and Bloom filter	Does not support scalability require high computational resource
<b>Incentive Distribution Techniques</b>				
[36]	Sharing of quality data among IoT devices for social welfare	A blockchain-based incentive driven data sharing scheme	Stackelberg game theoretic approach	Lack of fairness in incentive distribution
[37]	Integration of AIoT and FL to address the privacy issues	A iterative double-auction (IDA) mechanism to maximize social welfare	Reinforcement learning based double auction mechanism	Lack of fairness and require computational overhead
[38]	To incentivize server resources and ensure the quality of services	End-to-End service auction mechanisms for edge computing	COMputing Service Auction (COMSA)	Does not provide collaboration environment
[43]	To address the privacy and data freshness challenges in healthcare metaverse	Blockchain-based user-centric incentive mechanism approach for optimal data freshness	Age of Information (AoI)-based contract model under Prospect Theory (PT)	Require trust assessment to share data, does not support data exchange security
[44]	To allow Pareto improvements in supply chain members for collaboration	Blockchain-based three-level collaborative framework for supply chain	Shapley value	Require secure environment for data exchange, does not support data trustworthiness
[45]	To improve the trade-off between Pareto efficiency and distribution	Federated learning based incentive distribution mechanism	Shapley value and Pareto Optimality	Lack of distribution fairness, scalability issues
[46]	To achieve high accuracy and improve the efficiency of crowd-sourced system	Federated learning based enhanced incentive mechanism	Shapley value and contract theory	Require secure sharing mechanism, lacks verification of truthful data

**Our Contribution:** In this paper we proposed a secure collaborative data sharing framework using consortium blockchain technology. To control access to data we proposed an authentication protocol using ECC and HMAC to preserve the integrity of data and user anonymity. To ensure data truthfulness we employed an incentive distribution mechanism using the Shapley value that ensures fairness in profit distribution among data providers. In addition, consortium blockchain provides an efficient and secure environment for data sharing and storing of transactions on the ledger.

- *Associativity:* A group is associative if and only if it holds  $(a \circ b) \circ c = a \circ (b \circ c)$ ,  $\forall a, b, c \in G$ .
- *Identity:* There exists an element 1 called identity element such that  $a \circ 1 = 1 \circ a$ ,  $\forall a \in G$ .
- *Inverse:*  $\forall a \in G$  has an inverse element that exists in that group, such that  $a \circ a^{-1} = a^{-1} \circ a$ .

After describing the cyclic group properties, it is clear that one-way functions, the Discrete Logarithm Problem (DLP), can quickly be evaluated in a cyclic group..

*Definition 1 Discrete Logarithm Problem:* For a finite cyclic group  $Z_p^*$  with the order  $p - 1$  having a primitive element  $\alpha \in Z_p^*$ , and another element  $\beta \in Z_p^*$ . The DLP is to determine the element  $x$ , such that  $1 \leq x \leq p - 1$ .

$$\alpha^x \equiv \beta \pmod{p} \quad (1)$$

Roughly speaking,  $x$  must exist because  $\alpha$  is a primitive element that must have the power  $x$  that generates the element of the group. so  $x$  is said to be the discrete logarithm of  $\beta$  with

the base of  $\alpha$ . It can be denoted as:

$$x = \log_{\alpha} \beta \pmod{p} \quad (2)$$

It is tough to compute discrete logarithm problems when significant parameters are used. It needs extensive computation to solve the problem with the different attacks. In practice, DLP always considers in the cyclic group  $Z_p^*$  which is vulnerable to Pohling-Hellman attack. The cardinality of  $Z_p^*$  with the large prime number  $p$  is  $p - 1$  which is not a prime.

*Definition 2 Generalized Discrete Logarithm Problem:* For a finite cyclic group  $(G, \circ)$  of the order  $n$  and a primitive element  $\alpha$  having same order as a  $G$ . There is another element  $\beta \in G$ , DLP is to find a integer  $x$ , as  $1 \leq x \leq n$ , such that:

$$\beta = \alpha \circ \alpha \circ \alpha \circ \dots \circ \alpha = \alpha^x \quad (3)$$

### C. SHAPLEY VALUE

Game theory offers numerous profit distribution decision-making frameworks, including the Shapley value, which

evaluates each coalition data provider’s marginal contributions and equation to calculate participants’ contributions.

$$\varphi_i(N, w) = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(|N| - |S| - 1)!}{N!} [w(S \cup \{i\}) - w(S)] \quad (4)$$

The Shapley value helps distribute revenue fairly in collaborative atmospheres. Thus, our model uses the Shapley value to calculate revenue based on estimating resource use, fairness, and avoiding equalitarianism. To evaluate the performance certain requirements must be met as follows:

- **Efficiency:** The total worth of all participants equals the total money earned by the grand alliance. In other words, the sum of each participant’s Shapley value equals the group’s overall revenue, as shown by  $\sum_{i \in N} \varphi_i(w) = w(N)$ .
- **Symmetry:** If two coalition participants  $i$  and  $j$  make the same contribution, then their Shapley values should be the same s.t.,  $w(S \cup i) = w(S \cup j)$ .
- **Linearity:** If two coalitions have revenue functions  $v$  and  $w$ , the gain distribution derived from both coalitions’ worth functions should be equal to the sum of the gain distributions derived from each worth function individually, as in  $\varphi_i(v + w) = \varphi_i(v) + \varphi_i(w) \quad \forall i \in N$ .
- **Null Player:** The profit for a player  $i$  in a game is zero if the player’s contribution is zero,  $w(S \cup i) = w(S) \quad \forall i \notin S$ .

#### IV. PROBLEM DEFINITION AND SCOPE

##### A. SYSTEM MODEL

As shown in Figure.1, the proposed system model has participants, including data providers, data consumers, and blockchain authenticator.

- **Blockchain Authenticator (BA):** The blockchain authenticator is a critical component of our designed blockchain-based data-sharing system. It controls system communication and initializes when it receives a request from a user. It registers and authenticates the participants to restrict data access to authorized entities.
- **Data Provider (P):** Data providers hold dynamic industrial data. This data may be utilized for market research and industry monitoring. Data providers give data to consumers’ requests for incentives.
- **Data Consumer(C):** Data consumers utilize data to get insights or make optimal decisions. Data consumers request and get data from data providers via the blockchain in return for incentives.
- **Smart Contracts:** These self-executing, programmable contracts have coded terms and conditions. They reduce transaction costs, boost trust, and eliminate intermediaries. Business communication and transaction platforms use smart contracts. In traditional business models, payments are made indirectly through multi-tier models. However, in blockchain-based systems, payments are made directly through smart contracts.

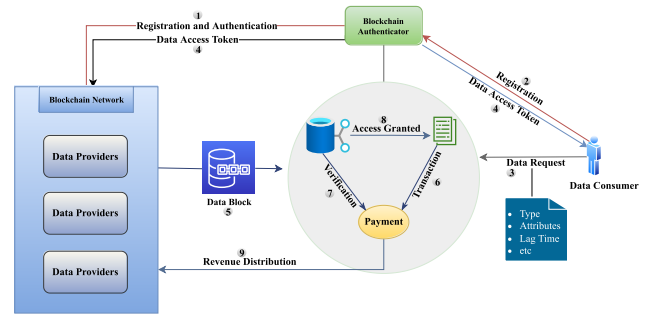


FIGURE 1. Proposed system model of data sharing system.

Registering with the blockchain authority secures participant communication. When users register, the blockchain broadcasts public parameters to the network. Both data providers and consumers need to register with the blockchain network to participate in data sharing. Consumers who have registered can send data requests to providers using the secure communication channel established by the blockchain.

##### B. THREAT MODEL

As per the Dolev and Yao [48] and CKadversary [49] threat models, our security model assumes an adversary capable of active and passive attacks. This adversary can intercept, modify, delete, and insert messages between entities. The attacker can also determine the communicating entities’ long-term private key for more complex attacks.

##### C. SECURITY REQUIREMENTS

For secure data transmission, security and privacy procedures are required. The following are the security requirements that we consider for our model:

- **Authenticity:** In a sharing system, participants can verify the identity of each other as data is shared with the consumer that may be malicious or legitimate and needs verification because the unauthorized access is a threat to the privacy of data. So, we considered this security requirement while designing the security protocol for the data-sharing system.
- **Impersonation Attack:** To impersonate a legitimate participant, the attacker needs to generate secret keys used for authentication and verification. However, an attacker cannot manipulate the communication between the parties as the intended receiver first verifies the sender’s authenticity using agreed parameters.
- **Replay Attack:** In this attack, the attacker pretends to be a legitimate participant and replaces the message sent by an actual legitimate party. The attacker reuses the parameters of the previous session to hijack the current session. Hence, our proposed protocol provides security against replay attacks.
- **Unlinkability Attack:** It is a privacy attack that discloses communication links between system participants are called unlinkability attacks. An attacker attempts

to disclose the anonymity of communication between them.

- **Traceability Attack:** An attempt by an attacker to trace back the communication flow between the participants to discover the identities of the communicating participants.
- **DDoS Attack:** As we have used blockchain technology, our protocol is resistant to DDoS attacks due to the decentralized nature of the network. The attacker may attack a single gateway that does not pose a threat to system performance in the presence of multiple gateways.
- **Length Extension Attack:** The attacker tries to calculate the hash of the message without knowing the actual message. The attacker generates the internal state by using the hash value. HMAC is resistant to this attack by truncating the hash value with SHA256/512. Hence, Our protocol prevents the system from this type of attack.

TABLE 2. Notations used in the proposed framework.

Notations	Description
$EC_{\mathbb{P}}$	Elliptic Curve of prime order
$\alpha$	Primitive element
$B_A$	Blockchain Authenticator
$\mathcal{P}_i, \mathcal{C}_i$	Data provider and consumer $i$ , respectively
$k_a, k_p$	Secret keys associated with $B_A$ and $\mathcal{P}$
$PK_a, PK_p$	Public keys of $B_A$ and $\mathcal{P}$
$\widehat{\mathcal{P}}_i, \widehat{\mathcal{C}}_i$	Assigned IDs to $\mathcal{P}_i$ and $\mathcal{C}_i$ , respectively by $B_A$
$\mathcal{R}_{\mathcal{E}_i}$	Data request from $\mathcal{C}_i$
$SK_{ap}, SK_{ac}$	Shared secret keys of $\mathcal{P}$ and $\mathcal{C}$ with $B_A$ , respectively
$m_a, m_{a1}$	Messages initiated by $B_A$
$m_c, m_{c1}$	Messages initiated by $\mathcal{C}$
$m_p$	Message initiated by $\mathcal{P}$
$R_a, R_c, R_p$	Random numbers of $B_A, \mathcal{C}$ , and $\mathcal{P}$ , respectively
$\zeta_i$	Transaction ID of $\mathcal{C}_i$
$\bar{T}_i$	Access Token generated for $\mathcal{C}_i$
$AT_{\mathcal{P}_i}, AT_{\mathcal{C}_i}$	Encrypted Tokens to $\mathcal{P}$ and $\mathcal{C}$ , respectively from $B_A$
$S$	Coalition of $\mathcal{P}_i$
$F_j$	Data value factors
$\mathcal{V}_{\mathcal{P}_i}$	Data value held by $\mathcal{P}_i$
$w(S)$	Worth function generated by $S$
$M_{\mathcal{P}_i}$	Marginal contribution of $\mathcal{P}_i$
$\phi(\mathcal{P}_i, S)$	Shapley value for $\mathcal{P}_i \subseteq S$
$\Re_{\mathcal{P}_i}$	Allocated revenue to each player

## V. DESCRIPTION OF THE PROPOSED SECURITY FRAMEWORK

In this section, we have briefly described our security mechanism for data-sharing systems.

Let  $\mathcal{P} = \{\mathcal{P}_1, \mathcal{P}_2, \mathcal{P}_3, \dots, \mathcal{P}_i\}$  represent the set of data providers, and  $\mathcal{C} = \{\mathcal{C}_1, \mathcal{C}_2, \mathcal{C}_3, \dots, \mathcal{C}_i\}$  represent the set of data consumers. Both  $\mathcal{P}_i$  and  $\mathcal{C}_i$  can participate in data sharing after registering. Our protocol comprises two phases as follows:

### A. REGISTRATION PHASE

$\mathcal{P}_i$  registers with  $B_A$  during this phase by sending a registration request. When  $B_A$  receives  $\mathcal{P}_i$ 's request, it generates registration parameters.  $B_A$  selects the secret key  $k_a \in Z_p^*$  and calculates public key  $PK_a = k_a \cdot \alpha$ , assigns new ID ( $\widehat{\mathcal{P}}_i$ ) to  $\mathcal{P}_i$ , such that  $\mathcal{P}_i \neq \widehat{\mathcal{P}}_i$ , and hashes it along with  $PK_a$  and  $T$ .

$$m_a = H(PK_a \parallel \widehat{\mathcal{P}}_i \parallel T) \quad (5)$$

$B_A$  sends  $(m_a, PK_a, EC_{\mathbb{P}}, \alpha)$  to  $\mathcal{P}_i$  to complete the registration process. When  $\mathcal{P}$  receives the parameters, it first verifies the message  $m_a$  by computing a hash and comparing it with a received message ( $m_a = m'_a$ ) under the assumption that  $H(\cdot)$  is publicly known.

$$m'_a = H(PK_a \parallel \widehat{\mathcal{P}}_i \parallel T) \quad (6)$$

Now,  $\mathcal{P}_i$  chooses the secret key  $k_p \in Z_p^*$  and computes its public key  $PK_p = k_p \cdot \alpha$ .  $\mathcal{P}_i$  transmits  $PK_p$  to  $B_A$  in order to compute the shared secret between  $B_A$  and  $\mathcal{P}$ . In addition, when  $B_A$  receives  $PK_p$  from  $\mathcal{P}$ , it computes the shared secret.

$$SK_{ap} = k_p \cdot PK_a \quad (7)$$

$$SK_{ap} = k_a \cdot PK_p \quad (8)$$

Both  $\mathcal{P}_i$  and  $B_A$  now have the same secret ( $SK_{ap}$ ) computed in Equations 7 and 8. When  $B_A$  shares the secret key ( $SK_{ap}$ ) with  $\mathcal{P}$ , it becomes eligible to join the blockchain network.

### Algorithm 1 Registration Phase

**Input :** Data provider  $\mathcal{P}_i$ , Authority  $B_A$

**Output:** Registered  $\mathcal{P}_i$ , Shared secret  $SK_{ap}$

$\mathcal{P}_i$  initiates registration by sending a request to  $B_A$ ;

$B_A$  generates registration parameters;;

- Selects secret key  $k_a \in Z_p^*$ ;

- Computes public key  $PK_a = k_a \cdot \alpha$ ;

- Assigns a new unique ID  $\widehat{\mathcal{P}}_i$  to  $\mathcal{P}_i$ ;

- Hashes  $\widehat{\mathcal{P}}_i$  along with  $PK_a$  and  $T$  to compute  $m_a$

as:  $m_a = H(PK_a \parallel \widehat{\mathcal{P}}_i \parallel T)$ ;

$B_A$  sends  $(m_a, PK_a, EC_{\mathbb{P}}, \alpha)$  to  $\mathcal{P}_i$  to complete the registration;

$\mathcal{P}_i$  verifies the received message by computing  $m'_a$ ;;

$m'_a = H(PK_a \parallel \widehat{\mathcal{P}}_i \parallel T)$ ;

**if**  $m_a = m'_a$  **then**

$\mathcal{P}_i$  selects secret key  $k_p \in Z_p^*$  and computes its public key  $PK_p = k_p \cdot \alpha$ ;

$\mathcal{P}_i$  transmits  $PK_p$  to  $B_A$  for shared secret computation;

$B_A$  computes the shared secret  $SK_{ap}$  as follows;;

$SK_{ap} = k_p \cdot PK_a$ ;

$SK_{ap} = k_a \cdot PK_p$ ;

$\mathcal{P}_i$  and  $B_A$  now possess the same secret  $SK_{ap}$ ;

Upon sharing  $SK_{ap}$  with  $\mathcal{P}_i$ , it becomes eligible to join the blockchain network;

**end if**

**else**

| Registration failed; abort the process;

**end if**

### B. AUTHENTICATION PHASE

$\mathcal{C}_i$  must register with  $B_A$  to obtain a data access token ( $AT$ ) to access the data from  $\mathcal{P}_i$ . Now  $\mathcal{C}_i \rightarrow (\mathcal{R}_{\mathcal{E}}, \theta^{\mathcal{P}_i}) \rightarrow B_A$ . After receiving  $\mathcal{R}_{\mathcal{E}}$ ,  $B_A$  checks to see if  $\mathcal{C}_i$  already exists in

the DL, if this is the case,  $B_A$  will end the process. Otherwise,  $B_A$  generates a random number or nonce  $N_a$  and  $SK_{ac}$ .

$$SK_{ac} = H(\widehat{C}_i \parallel N_a \parallel k_a) \quad (9)$$

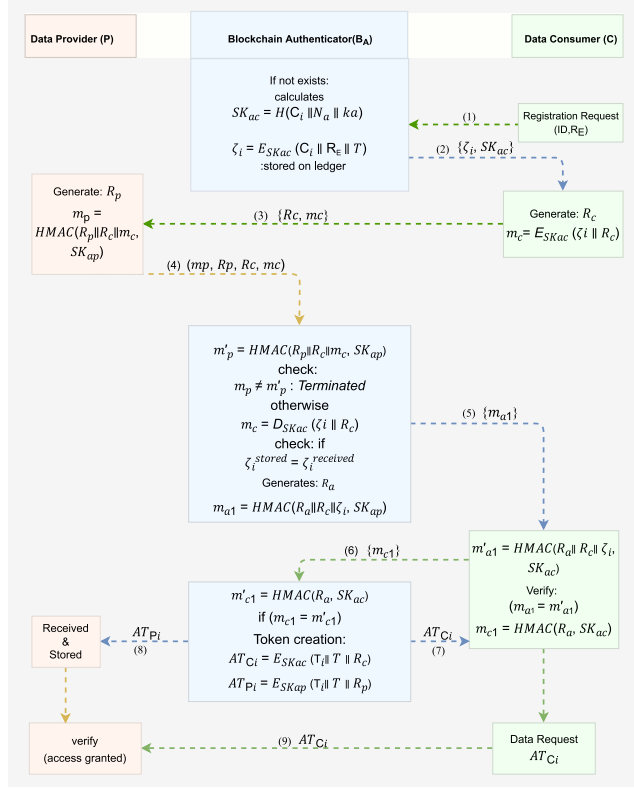


FIGURE 2. Authentication and token generation phase.

Secret key  $k_a \ni B_A$  and is utilized to compute the shared secret key  $SK_{ac}$  with  $C_i$ .  $B_A$  encrypts the requested  $C_i$ 's ID ( $\widehat{C}_i$ ),  $\mathcal{R}_E$ , and  $T$  with  $SK_{ac}$  and stores this encrypted message as transaction ID ( $\zeta_i$ ) on the blockchain ledger for verification.

$$\zeta_i = E_{SK_{ac}}(\widehat{C}_i \parallel \mathcal{R}_E \parallel T) \quad (10)$$

Now,  $\zeta_i$  and  $SK_{ac}$  are shared with  $C_i$  via a secure connection to begin the process of obtaining the token.  $C$  encrypts  $m_c = E_{SK_{ac}}(\zeta_i \parallel R_c)$  and sends this message  $m_c$  to  $P_i$ . After receiving the message  $m_c$  from  $C_i$ , it generates  $R_p$  and then uses  $SK_{ap}$  to calculate the HMAC of the message.

$$m_p = HMAC(R_p \parallel R_c \parallel m_c, SK_{ap}) \quad (11)$$

$P_i : m_p, R_p, R_c, m_c \rightarrow B_A$ , if  $B_A : m_p \equiv m'_p$ . Then  $B_A$  verifies  $\zeta_i$  by decrypting  $m_c : K_{ac}$  and comparing it to  $\zeta_i$  stored in DL. If  $\zeta_i^{stored} = \zeta_i^{received}$ , the loop continues.

$$m'_p = HMAC(R_p \parallel R_c \parallel m_c, SK_{ap}) \quad (12)$$

After verifying the Transaction ID ( $\zeta_i$ ),  $B_A$  computes  $m_{a1} = HMAC(R_a \parallel R_c \parallel \zeta_i, SK_{ap})$  sends  $m_{a1} \rightarrow C_i$ .  $\zeta_i$  represents the transaction ID associated with  $C_i$ 's request. When  $m_{a1}$  is received,  $C_i$  verifies  $m_{a1} \equiv m'_{a1} =$

$HMAC(R_a \parallel R_c \parallel \zeta_i, SK_{ac})$ . If yes, then  $C_i$  uses  $SK_{ac}$  to determine  $m_{c1} = HMAC(R_a, SK_{ac})$  of received  $R_a$  and sends  $m_{c1} \rightarrow B_A$ .  $B_A$  verifies  $m_{c1} \equiv m'_{c1} = HMAC(R_a, SK_{ac})$  after receiving it from  $C_i$ . If  $m_{c1} \equiv m'_{c1}$ ,  $B_A$  produces token  $T_i$ ,  $T$ , encrypts with  $SK_{ac} \rightarrow C_i$ . Also, compute it for  $P_i$  encrypts with  $SK_{ap}$ .

$$AT_{C_i} = E_{SK_{ac}}(T_i \parallel T \parallel R_c) \quad (13)$$

$$AT_{P_i} = E_{SK_{ap}}(T_i \parallel T \parallel R_p) \quad (14)$$

## VI. PROFIT AND REVENUE ASSESSMENT STRATEGIES

### A. PROFITABILITY FRAMEWORK

Let  $\mathcal{C} = \{C_i \mid i = 1, \dots, C\}$  be the set data consumers submits the request  $\mathcal{R}_E = \{r_j \mid j = 1, \dots, R\}$  to data providers  $\mathcal{P} = \{P_i \mid i = 1, \dots, P\}$  such that  $C_i = r_j$  and  $r_j = (d_{ij}t, AT_{C_i}) \rightarrow P_i$ . Then,  $\mathcal{P}$  form a coalition  $S$  as,  $P_i \oplus S \subseteq \mathcal{P} \wedge |S| \leq |\mathcal{P}|$ .  $d_{ij}$  is the requested data type and  $t$  is the life span. Each  $P_i \subseteq S \iff d_{ij}^{A_i}$ , here  $A_i$  data held by  $P_i \in \mathcal{P}$ . Let  $\mathbf{A}$  be a set of data characteristics such that  $P_i = \mathbf{A}_i$ . Then  $P_i$  is with  $\mathbf{A} = \{a_1, a_2, \dots, a_n\}$ , where  $a_n$  is data held by  $P_i$  number of attributes. We can determine the value of data by defining data uniqueness, quantity, and quality. By aggregating these factors we can define data value.

#### 1) UNIQUENESS (U)

Lets  $U_{P_i}$  is data uniqueness held by  $P_i$  and we can define a function that quantifies  $U_{P_i} = f(c_p, d_p^{an}, c_r)$  where  $c_p$ ,  $d_p^{an}$ , and  $c_r$  are frequency of occurrence, distinct attribute of data, and alignment of data with existing data, respectively. If the value of  $f$  is high such as,  $f(c_p, d_p^{an}, c_r) \iff d_p, d_p^{an} > c_p \wedge c_r = 1$ , then  $U_{P_i}$ :

$$U_{P_i} = \beta \cdot c_p - \gamma \cdot d_p^{an} + \delta \cdot \sum_{i=1}^n \cdot c_r^i \quad (15)$$

Here,  $\beta$ ,  $\gamma$  and  $\delta$  are weighting factors that define the relative importance of  $c_p$ ,  $d_p$ , and  $c_r$  respectively.  $\beta \cdot c_p$  denotes the frequency of occurrence which mean if  $c_p > d_p^{an}$  then  $U_{P_i}$  decreases because data is more prevalent. However,  $-\gamma \cdot d_p^{an}$  indicates that if its increase offsets the value of  $c_p$ . This trade-off recognizes that uniqueness is not exclusively defined by  $c_p$  or  $d_p^{an}$  but by their relative significance and balance.

#### 2) DATA QUALITY ( $\mathcal{Q}_i$ )

The quality of data held by  $P_i$  depends on its dimensions such as accuracy ( $\hat{A}$ ), life span ( $t$ ), validity ( $v$ ), and uniqueness ( $U_{P_i}$ ).  $\hat{A} \in [0, 1]$  is accuracy score for  $d_{ij}$  such that  $0 \leq \hat{A} \leq 1$ . Similarly, we can determine the life span or timeliness of data by mapping the time difference  $\Delta t$ . Suppose,  $t$  is the total score, then the linear mapping is:

$$t = g(\Delta t) = 1 - \left(\frac{\Delta t}{T}\right) \quad (16)$$

Here,  $T$  represents the maximum acceptable threshold of  $\Delta t$ . If the value of  $g(\Delta t) \simeq T$  means smaller the  $\Delta t$ ,

**Algorithm 2** Authentication and Token Generation Phase

**Input** : Data consumer  $C_i$ , Authority  $B_A$   
**Output**: Data access token  $AT_{C_i}$ , Data access token  $AT_{P_i}$

$C_i$  initiates registration with  $B_A$  for obtaining a data access token ( $AT$ ) to access data from  $P_i$ ;  
 $B_A$  checks the existence of  $C_i$  in the DL;  
**if**  $C_i$  does not exist in the DL **then**  
      $B_A$  generates a nonce  $N_a$  and a secret key  $SK_{ac}$ ;  
      $SK_{ac} = H(\widehat{C}_i \parallel N_a \parallel k_a)$ ;  
      $\zeta_i = E_{SK_{ac}}(\widehat{C}_i \parallel \mathcal{R}_E \parallel T)$ ;  
      $C_i$  receives  $\zeta_i$  and  $SK_{ac}$  via a secure connection;  
      $C_i$  encrypts  $m_c = E_{SK_{ac}}(\zeta_i \parallel R_c)$  and sends it to  $P_i$ ;  
      $P_i$  receives  $m_c$  from  $C_i$  and computes  
          $m_p = HMAC(R_p \parallel R_c \parallel m_c, SK_{ap})$ ;  
      $P_i : m_p, R_p, R_c, m_c \rightarrow B_A$ ;  
     **if**  $B_A$  verifies  $m_p \equiv m'_p$  **then**  
          $B_A$  decrypts  $m_c$  using  $SK_{ac}$  to obtain  $\zeta_i$ ;  
         **if**  $\zeta_i^{stored} = \zeta_i^{received}$  **then**  
              $B_A$  computes  
                  $m_{a1} = HMAC(R_a \parallel R_c \parallel \zeta_i, SK_{ap})$  and  
                 sends  $m_{a1}$  to  $C_i$ ;  
              $C_i$  verifies  
                  $m_{a1} \equiv m'_{a1} = HMAC(R_a \parallel R_c \parallel \zeta_i, SK_{ac})$ ;  
             **if**  $m_{a1} \equiv m'_{a1}$  **then**  
                  $C_i$  computes  $m_{c1} = HMAC(R_a, SK_{ac})$   
                 and sends  $m_{c1}$  to  $B_A$ ;  
                  $B_A$  verifies  
                      $m_{c1} \equiv m'_{c1} = HMAC(R_a, SK_{ac})$ ;  
                 **if**  $m_{c1} \equiv m'_{c1}$  **then**  
                      $B_A$  generates token  $T_i$  and encrypts  
                     it with  $SK_{ac} \rightarrow C_i$ ;  
                      $P_i$  encrypts  $T_i, T$ , and  $R_p$  with  
                      $SK_{ap}$ ;  
                      $AT_{C_i} = E_{SK_{ac}}(T_i \parallel T \parallel R_c)$ ;  
                      $AT_{P_i} = E_{SK_{ap}}(T_i \parallel T \parallel R_p)$ ;  
                     **end if**  
                 **end if**  
             **end if**  
         **end if**  
     **end if**  
**end if**

therefore, greater the  $t$ . If  $g(\Delta t) > T$  then  $d_{ij}$  has low  $t$ . For example, we have  $d_{ij}^o$  having the  $\Delta t = 12h$  and we have set  $T = 24h$ . Now the value of  $t = 1 - \frac{12}{24} = 1 - 0.5 = 0.5$ , which means that the  $d_{ij}$  has 50% by considering  $T$  24 hours. By aggregating the values  $F = U_{P_i} + \hat{A} + t + v$ , we can define the  $Q_i$  as:

$$Q_i = \sum_{i=1}^n \sum_{j=1}^m w_i \cdot F_j \quad \forall i, j = 1, 2, 3 \dots n \quad (17)$$

Here,  $w_i$  denotes the relative importance of each dimension for an overall assessment of data quality. Now we can

calculate data value  $V_{P_i}$  held by  $P_i$  by aggregating these factors.

$$V_{P_i} = \rho \cdot \sum_{i=1}^P (U_{P_i} + Q_i + v) \quad (18)$$

The value  $\rho$  is defined based on importance of  $d_{ij}$  by  $P_i$  based  $C_i$  preferences. So,  $S_{V_{P_i}} \subseteq \mathcal{P}$  generates  $w(S)$ :

$$w(S) = \sum_{i \in S}^n V_{P_i}(S) \quad (19)$$

**B. REVENUE ESTIMATION METHODOLOGY**

After determining the worth function  $w(S)$  for coalition  $S$ , we can calculate the incentive of each  $P_i \in S$  upon joining the coalition  $S$ . For each  $P_i$  its marginal contribution  $m_c$  is determined for all possible permutations  $\theta$ . In other words,  $M_{P_i}$  is a difference of worth generated by  $S$  when  $P_i$  joined or when absent from the  $S$ . It can be calculated as:

$$M_\theta(P_i) = w(S \cup P_i) - w(S) \quad (20)$$

where  $w(S \cup P_i)$  denotes the worth function of  $S$  when  $P_i \in S$ , and  $w(S)$  when  $P_i \notin S$ . The shapely value  $\phi(P_i, S), \forall P_i \in S$  across  $\theta$  by averaging  $M_\theta(P_i)$ .

$$\phi(P_i, S) = \frac{1}{N!} \sum_{\theta \in \Theta} M_\theta(P_i) \quad (21)$$

where  $N$  is the total number of players in  $S$ ,  $\Theta$  is the set of all permutations of  $P_i \in S$ , and  $M_\theta(P_i)$  represents the marginal contribution of  $P_i$  in a specific permutation  $\theta$ . By normalization of  $\phi(P_i, S)$  it ensures that profit is fairly distributed among  $P_i \in S$  such as:

$$\bar{\phi}(P_i, S) = \frac{\phi(P_i, S)}{\sum_{P_i \in S} \phi(P_i, S)} \quad (22)$$

Based on  $\bar{\phi}(P_i, S)$  for each  $P_i$  we can determine the total incentive or revenue  $\mathfrak{R}_{P_i}$  allocated to each player as:

$$\mathfrak{R}_{P_i} = \bar{\phi}(P_i, S) \times w(S) \quad (23)$$

**C. REVENUE DISTRIBUTION AMONG DATA PROVIDES**

To determine the contribution of each  $P_i$ , then we have to measure the change in the marginal contribution of data providers of the coalition. Therefore, to calculate this factor we measure the performance of the sharing system as that evaluates the worth function in Equation 4. To measure the performance, we will use the F1-score as it can evaluate the performance of the model more efficiently in various scenarios.

It should be noted that a model with a high F1-score is considered better in performance. The contribution of data providers is based on the impact of their shared data consequently affecting revenue.



Here we assume that each  $\mathcal{P}_i \in \mathcal{P}$  have the same type of data in a coalition such that,  $\mathcal{R}_{\mathcal{E}} = D_i$  that generates the revenue  $w(\mathcal{P})$ . The contribution of data providers is considered as an impact factor assuming that all the data providers share data with a positive impact on the system for the legality of Shapley value for  $\mathcal{P}_i$ . We evaluate the F1-score for data shared by coalition data providers.

The relation between F1-score and Shapley value for a game  $(\mathcal{P}, w)$  is described as:

$$\varphi_i(\mathcal{P}, w) = \sum_{S \in \mathcal{P}} w(|S|) \frac{F1(S \cup \{i\}) - F1(S)}{F1(\mathcal{P})} \quad (24)$$

Equation 24  $w(|S|)$  is the weighted factor as described in Equation 19,  $F1(S)$  shows the F1-score produced by coalition  $(S \subseteq \mathcal{P})$   $S$  of  $\mathcal{P}$ . So, from above equations we can determine the  $f_i(\mathcal{P}) = \sum_{S \in \mathcal{P}} w(|S|) \frac{F1(S \cup \{i\}) - F1(S)}{F1(\mathcal{P})}$  is participants  $i$  impact factor to the coalition  $S$ . Also the aggregated Shapley value for participants  $i$  can be defined as:

$$\phi_i(\mathcal{P}) = \sum_{\mathcal{P}_i \in \mathcal{P}} \varphi_i(w) \quad (25)$$

Above mention equation satisfies the property of *efficiency* meaning that the sum of all revenue generated by participants is equal to the grand coalition as:

$$\varphi_{\mathcal{P}_1}(w) + \varphi_{\mathcal{P}_2}(w) + \varphi_{\mathcal{P}_3}(w) + \dots + \varphi_{\mathcal{P}_i}(w) = \phi_{\mathcal{P}}(w) \quad (26)$$

As from Equation 20 we can determine the marginal contribution of participant  $i$ . According to this intuition, we can also calculate the marginal contribution of participants  $i$  in F1-score such as:

$$\Delta_i(F1, S) = F1(S \cup \{i\}) - F1(S) \quad (27)$$

To satisfy Equation 4 we have to normalize the marginal contribution of participant  $i$  to the F1-score. We need to divide  $\Delta_i(F1, S) = F1(S \cup \{i\}) - F1(S)$  with  $F1(\mathcal{P})$  to hold the Equation 19, i.e.,  $\phi(\mathcal{P}) = \varphi_i(\mathcal{P})$ . Here it is important to mention that  $F1(\mathcal{P})$  is the maximum F1-score by the coalition.  $f_i(\mathcal{P})$  is the impact factor of participant  $i$  that eventually used to determine the participants percentage contribution to F1-score. In a considered data-sharing system  $\mathcal{P}_i$  generates the revenue  $\varphi(\mathcal{P})$  by sharing data. The Shapley value to distribute the revenue among the  $\mathcal{P}_i$ , then the Shapley value for data providers  $\phi_{\mathcal{P}} = \sum_{\mathcal{P}_i \in \mathcal{P}} \varphi_{\mathcal{P}_i}$  is define in this way.

Here we assume that,  $\mathcal{P}$  is sharing same type of data and  $S_{\mathcal{P}}^i = \{\mathcal{P}' \mid \mathcal{P}' \subseteq \mathcal{P} / \{\mathcal{P}_i\}\}$  is a set of data providers  $\mathcal{P}$  excluding the player  $\mathcal{P}_i$ . According to the F1-score, we can define the revenue distribution for a set of players using the Shapley value:

$$\varphi_{\mathcal{P}_i}(\mathcal{P}, w) = \varphi_{\mathcal{P}_i}^i(\mathcal{P})w(\mathcal{P}) \quad (28)$$

Equation 28 shows that revenue of each participant in a coalitional game, i.e.,  $\varphi_{\mathcal{P}}^i = \{\varphi_{\mathcal{P}}^1, \varphi_{\mathcal{P}}^2, \dots, \varphi_{\mathcal{P}}^{|\mathcal{P}|}\}$ . So, we can define  $\varphi_{\mathcal{P}}^i$  for each  $\mathcal{P}_i \in \mathcal{P}$  in term of F1-score as follow:

$$\varphi_{\mathcal{P}}^i(\mathcal{P}) = \sum_{\mathcal{P}' \in S_{\mathcal{P}}^i} w(|\mathcal{P}'|) \frac{\Delta_{\mathcal{P}_i}(F1, \mathcal{P}')}{F1(\mathcal{P})} \quad (29)$$

Normalizing the equation 29 by dividing it with  $F1(\mathcal{P})$  to receive percentage of  $\mathcal{P}_i$  contribution to F1-score. Moreover,  $\varphi_{\mathcal{P}}^i$  is not the same for the data providers because of their contribution. Hence, the aggregated Shapley value shows that the sum of all contributions is equal to the grand coalition contribution, i.e.,  $\phi_{\mathcal{P}} = w_{\mathcal{P}}^i(S)$ .

---

**Algorithm 3** Data Sharing System Algorithm

---

**Input** : Request RE from data consumers

**Output:** Data block and revenue distribution among data providers

Initialize blockchain authenticator and data providers;

**for each data provider P do**

    P registers with blockchain authenticator;

**end for**

**for each data consumer C do**

    C registers with blockchain authenticator and obtains data access token;

    C sends request RE for desired data;

**for each data provider P do**

**if P has data with attributes matching RE then**

            P publishes attributes of held data and monetary value;

            Coalition S forms with all data providers having matching data attributes;

**if all P in S verify commitment from C then**

                Data block is generated;

                C pays incentive to access data block;

                Revenue is generated and distributed among P according to their contribution;

**end if**

**end if**

**end for**

**end for**

---

**VII. EVALUATION OF SECURITY AND PERFORMANCE**

In this section, we performed a security evaluation of our proposed security protocol using formal and informal methods.

**A. FORMAL SECURITY VERIFICATION USING ROR MODEL**

This section describes an approach to formally validating the security of the data access token using the ROR mathematical model proposed by Abdalla et al. [50]. This proposed protocol involves three participants:  $\mathcal{P}$ ,  $\mathcal{C}$ , and  $B_A$ . Let us denote  $\mathcal{P}^r$  and  $\mathcal{C}^s$  as examples that correspond to  $r$  and  $s$ , respectively. The adversary, denoted as **A**, commences targeted inquiries as a component of the adversarial activity.

*Theorem 1:* Assume that the adversary **A** intends to obtain the token  $AT$  within a feasible period of time. The adversary's advantage  $Ad_A$  is constrained by the following expression  $Ad_A \leq \frac{q_H^2}{|F|} + 2Ad_A^{ECDDH}$ . Here,  $q_H$ ,  $|F|$ , and  $Ad^{ECDDH}$

represent the hash query, the range for the hash function  $H(\cdot)$ , and the adversary  $A$ 's advantage in solving the ECDLP problem, respectively.

**Proof:** We perform the proof of our proposed protocol to verify the security of the access token ( $AT$ ). To demonstrate we consider three games such as  $G_i$ , where  $i \in [0, 2]$  and event  $E_{AG_i}$  which can be described as it can predict the random bit  $c$  and  $Pr[E_{AG_i}]$  describe as competitive advantage.

**Game ( $G_0$ ):** This game is designed to launch a real-time attack, with the initial selection of bits  $c$  being chosen at random. Therefore, semantic analysis can be used to derive insights.

$$Ad_A = [2Pr[E_{AG_i}] - 1] \quad (30)$$

This game is designed to launch a real-time attack, with the initial selection of bits  $c$  being chosen at random. Therefore, semantic analysis can be used to derive insights.

**Game ( $G_1$ ):** In this game  $A$  intercepts the message that is being exchanged between participants by running Execute query. Then it executes the REVEAL and TEST queries to verify the correctness of  $AT$  generated and shared between  $\mathcal{P}$  and  $\mathcal{C}$ . However, the  $AT$  is comprised of long secret and random numbers that are not known to the attacker. Therefore, intercepting the communication will not disclose the  $AT$ , and eventually, it provides equivalence between  $G_0$  and  $G_1$  winning probability.

$$Pr[E_{AG_1}] = Pr[E_{AG_0}] \quad (31)$$

**Game ( $G_2$ ):** In this scenario, the attacker formulates an active attack query by executing a HASH query. As the messages sent between all participants,  $\mathcal{P}$ ,  $\mathcal{C}$ , and  $B_A$ , are either hashed or encrypted, the attacker is unable to compromise the confidentiality of  $AT$ . In addition, the protocol utilizes three randomized integers, namely  $R_c$ ,  $R_p$ , and  $R_a$ , which makes it impossible for the attacker to determine them because of the ECDLP. To obtain  $AT$ , an attacker must computationally determine  $R_c$ ,  $R_p$ , and  $R_a$ , and then carry out the hash query to identify a collision. Therefore,  $G_2$  is identical to  $G_1$  in terms of winning probability, except for the hash collision. Therefore, by merging the ECDLP with the birthday paradox, the subsequent conditions arise:

$$Pr[E_{AG_1}] - Pr[E_{AG_2}] \leq Ad_{A^{ECDLP}} + \frac{(q_H^2)}{2F} \quad (32)$$

Upon completing all the games,  $A$  must possess the ability to empirically determine the accuracy of  $c$  bits to derive the inference of  $AT$  from it. Based on these facts, we have:

$$Pr[E_{AG_2}] = \frac{1}{2} \quad (33)$$

from equations (1) (2) and (4), we can obtain

$$Ad_A = |2Pr[E_{AG_2}] - 1| = Pr[E_{AG_1}] - Pr[E_{AG_2}] \quad (34)$$

Thus, this demonstration shows that an attacker will be able to ascertain the session key in polynomial time.

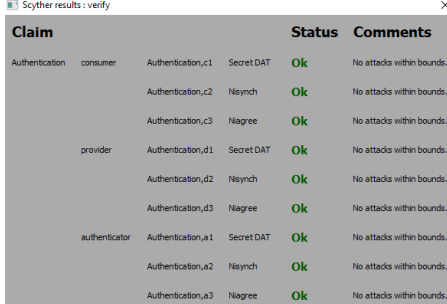
## B. FORMAL VERIFICATION OF SECURITY

Python-integrated Scyther Tool protocol assertions verified a security protocol. Scyther claims specify protocol security as shown in Table 3 including Secret, Nisynch, and Niagree claims. Our protocol specified consumer, provider, and authenticator roles utilizing Scyther's Security Protocol Description Language (SPDL).

TABLE 3. Considered claim events for proposed protocol.

Security Claims	Description
Secrecy	The information is sent using an untrusted communication channel, the adversary cannot reveal original information. Role executes this event in role specification.
Nisynch	This claim (R, Nisynch', R') ensures that the message sent from a sender is received by the intended receiver.
Niagree	It is a form of authentication that is based on an agreement between sender and receiver on data exchange that remains unchanged during the communication.

The claim events for which the proposed protocol is analyzed and its verification result is shown in Figure 3.



Claim	Status	Comments
Authentication consumer Authentication,c1 Secret DAT	Ok	No attacks within bounds.
Authentication,c2 Nisynch	Ok	No attacks within bounds.
Authentication,c3 Niagree	Ok	No attacks within bounds.
provider Authentication,d1 Secret DAT	Ok	No attacks within bounds.
Authentication,d2 Nisynch	Ok	No attacks within bounds.
Authentication,d3 Niagree	Ok	No attacks within bounds.
authenticator Authentication,a1 Secret DAT	Ok	No attacks within bounds.
Authentication,a2 Nisynch	Ok	No attacks within bounds.
Authentication,a3 Niagree	Ok	No attacks within bounds.

FIGURE 3. Security analysis results of proposed protocol.

The widely used simulation tool "AVISPA" will be used to verify the formal security of our proposed protocol. The protocol is written in High-Level Protocol Specification Language (HLPSL) for AVISPA. Four models can verify protocol security: "Constraint Logic-based Attack Searcher (CLAtSe)," "SAT-based Model Checker (SATMC)," "Tree Automata based on Automatic Approximations for Analysis of Security Protocol (TA4SP)," and "On-the-Fly Model Checker (OFMC)." Figure 4 shows that our proposed protocol is safe and resistant to adversarial attacks.

## C. INTUITIVE SECURITY ANALYSIS

In this section, our security protocol is evaluated theoretically against various malicious threats.

**Replay Attack:** When the  $k_a$ ,  $k_p$ , and  $AT_{\mathcal{P}_i}$ ,  $AT_{\mathcal{C}_i}$  are generated, both  $\mathcal{P}_i$ ,  $\mathcal{C}_i$  use predefined parameters. The association of  $SK_{ap}$ ,  $SK_{ac}$  and  $R_c, R_p, R_a$  in the exchange of messages between parties differs for each session. Hence, the session between parties is prevented from being hijacked.

**Secret Disclosure Attack:** As token ( $AT_{\mathcal{C}_i}$ ,  $AT_{\mathcal{P}_i}$ ) is encrypted using ( $SK_{ap}$ ,  $SK_{ac}$ ). An attacker cannot guess the keys generated by  $B_A$ . If the attacker managed to trace the participants but cannot disclose  $AT_{\mathcal{C}_i}$ ,  $AT_{\mathcal{P}_i}$  as it does not

```

SPAN 1.6 - Protocol Verification : Data_Sharing
File
% OFMC
% Version of 2006/02/13
SUMMARY
SAFE
DETAILS
BOUNDED_NUMBER_OF_SESSIONS
PROTOCOL
/home/span/span/testsuite/results/Data_Sharing_Protocol.if
GOAL
as_specified
BACKEND
OFMC
COMMENTS
STATISTICS
parseTime: 0.00s
searchTime: 0.00s
visitedNodes: 16 nodes
depth: 4 plies

```

FIGURE 4. AVISPA security analysis results of proposed protocol.

know the  $(SK_{ap}, SK_{ac})$ . Also,  $\mathcal{P}_i, \mathcal{C}_i, B_A$  use these keys to hash the messages  $(m_c, m_p, m_{a1}, m_{c1})$  by generating HMAC.

**Unlinkability Attack:** In our protocol, each  $\mathcal{C}_i$  has separate shared secret with  $B_A$  to generate  $T_i$  for each  $\mathcal{C}_i$ . If one of  $\mathcal{C}_i$ 's  $T_i$  is compromised, the other will not be threatened by this vulnerability. So, our protocol provides security against this attack by verifying  $T_i$  as  $B_A$  sends  $(AT_{\mathcal{P}_i}, AT_{\mathcal{C}_i})$  encrypted with  $(SK_{ap}, SK_{ac})$ .  $\mathcal{P}$  first verify the  $T_i$  and calculate the  $T$  if verified then grant access to data, otherwise discard the  $T_i$ .

**Traceability Attack:** Given that random numbers  $R_c, R_p$ , and  $R_a$  are used during message sharing and that these numbers are generated new for each session, an adversary cannot identify a fixed value to link parties. In addition, the adversary is unable to access any information during the current session because the parameters for each session are unique.

**Authenticity of Message:** Using HMAC provides an extra security layer during communication. Both  $\mathcal{P}$  and  $\mathcal{C}$  share the symmetric i.e.,  $SK_{ap}, SK_{ac}$  with  $B_A$  and send  $m_p = HMAC(R_p || R_c || m_c, SK_{ap})$  and  $m_{c1} = HMAC(R_a, SK_{ac})$  to  $B_A$ , respectively. Also,  $B_A$  sends  $m_{a1} = HMAC(R_a || R_c || \zeta_i, SK_{ac})$  to  $\mathcal{P}_i$ . If an attacker can capture messages  $(m_p, m_{c1}, m_{a1})$  but can not disclose the information as HMAC is verified using symmetric keys. Moreover, participants can verify  $\Leftrightarrow HMAC_{B_A}^{SK_{ap}} \simeq HMAC_{B_A}^{SK_{ac}}, \exists \forall \mathcal{P}_i || \mathcal{C}_i$ .

**DDoS Attack:** We have used a blockchain platform instead of a centralized platform reduces the probability of attack.  $\mathcal{P}_i$  and  $\mathcal{C}_i$  are distributed to  $B_A$ 's that minimize the attack surface. An attacker may target a single  $B_A$  that does not affect the others. That eliminates the changing secret  $(SK_{ac}, SK_{ap})$  for each session and prevents system desynchronization.

**Impersonating Attack:** The identity of legitimate participants is forged to use for unauthorized data access. If the attacker manages to acquire the secret key of  $\mathcal{C}_i$  during the registration process. Random numbers  $(R_c, R_a)$  are involved for each message preventing attackers from using the same key for a future session. Participant can verify message

integrity through nonce, such as  $SK_{ac} = H(\widehat{\mathcal{C}}_i || N_a || k_a)$ . Also messages  $(m_c, m_p, m_{a1}, m_{c1})$  are hashed before it is delivered.

**Length Extension Attack:** In this attack, an attacker uses the hash of message  $M_1$  e.g.,  $H(message || secret)$  to reveal the length of the message and modify it to  $M'_1$ . MD5 or SHA-1 is vulnerable to this attack until the SHA512-256 variant is proposed that calculates 512-size output and truncates it with a 256-bit extension. Such as,  $HMAC(SK_{ap}, m_p) = H((SK_{ap} \oplus opad) || H((SK_{ap} \oplus ipad) || m_p))$ . If the attacker can know the inner hashed message digest without using secret key  $SK_{ap}$  it can not calculate the outer fixed length digest  $(SK_{ap} || inner - hash)$ .

#### D. COMPARATIVE PERFORMANCE ANALYSIS

We have assessed the performance of our proposed security protocol for secure data sharing among system participants in terms of communication and computational cost. Our analysis was conducted on a laptop with an 11th Gen Intel(R) Core(TM) i7-1165G7 @ 2.80GHz 2.70 GHz processor, 12.0 GB RAM, and Windows 11 Home operating system.

To implement our proposed protocol, we have used PyCharm 2022.1.2 (Community Edition) and Python cryptographic library called "cryptography" by importing its functions such as "Fernet" (such as: `from cryptography.fernet import Fernet`) to use encrypt/decrypt functions. For hashing, we have used the "hashlib" library and cryptographic function HMAC of cryptography such as `from cryptography.fernet import HMAC`. To record the time for cryptographic operations we have used the python `time` library.

#### 1) COMPUTATIONAL COMPLEXITY ANALYSIS

Computational cost refers to the time taken by the protocol to execute its predefined operations. To evaluate the computation cost for the registration, authentication, and token generation process we have calculated the total execution time required to perform some significant cryptographic primitives involved during protocol implementation. The computation time to perform the point multiplication on an elliptic curve is  $T_m$ , one-time hashing and HMAC operations are  $T_h$ , the computational time for symmetric and asymmetric encryption/decryption is  $T_e$ , and the HMAC verification time is  $T_v$ .

In our protocol, there are three point-multiplication function operations, hashing three times, four times encryption/decryption, and four-time HMAC generation and verification. The data provider  $\mathcal{P}_i$  performs the point multiplication function  $(PK_a = k_a \cdot \alpha)$  during the initialization of the registration process. After receiving parameters from  $B_A$  it calculates the shared secret  $(SK_{ap} = k_a \cdot PK_p)$  and the calculation on  $B_A$  is  $(SK_{ap} = k_p \cdot PK_a)$ . Again these operations are performed during the registration process of data consumer  $\mathcal{C}_i$ . During the authentication and token generation process  $\mathcal{P}_i$  perform hashing operation  $m_a = H(PK_a || \widehat{\mathcal{P}}_i || T)$  to agree on shared secret with  $B_A$ , and

TABLE 4. Comparison of security features.

Security Features	[25]	[18]	[26]	[19]	[20]	[21]	[22]	[23]	[31]	[32]	Ours
Mutual Authentication	X	X	✓	✓	✓	✓	✓	✓	✓	✓	✓
Forward Secrecy	X	X	✓	✓	X	X	X	X	✓	✓	✓
Resilient to Impersonation Attack	✓	✓	X	✓	✓	✓	✓	✓	✓	✓	✓
Resilient to Unlinkability Attack	X	X	✓	X	✓	✓	X	X	✓	X	✓
Resilient to Traceability Attack	X	✓	X	X	X	✓	X	✓	X	X	✓
Resilient to Replay Attack	✓	✓	X	✓	✓	X	X	✓	✓	✓	✓
Resilient to Secret Disclosure Attack	✓	X	X	X	✓	✓	X	✓	✓	✓	✓
Resilient to DoS and DDoS	X	✓	X	X	✓	X	✓	X	X	✓	✓
Resilient to Length Extension	X	X	X	X	X	X	X	X	X	✓	✓

TABLE 5. Computation cost comparison.

Ref	Platform	Total Computational Cost	Approx Time
[18]	Blockchain	$2T_h + 5T_m + 4T_e + 2T_v$	1.92 ms
[19]	Non-Blockchain	$14T_h + 3T_m + 9T_e$	3.33 ms
[20]	Blockchain	$8T_h + 3T_m + 5T_e + 3T_v$	2.15 ms
[21]	Blockchain	$3T_h + 7T_m + 6T_e + 3T_v$	2.80 ms
[22]	Non-Blockchain	$4T_h + 8T_m + 6T_e$	2.84 ms
[23]	Non-Blockchain	$9T_h + 5T_m + 5T_e$	2.32 ms
[24]	Non-Blockchain	$13T_h + 2T_m + 8T_e$	2.89 ms
[31]	Non-Blockchain	$12T_h + 8T_m + 6T_e$	3.11 ms
[34]	Non-Blockchain	$5T_h + 2T_m + 9T_e$	2.93 ms
Ours	Blockchain	$3T_h + 3T_m + 4T_e + 4T_v$	1.64 ms

Constants:  $T_h = 0.03$  ms,  $T_m = 0.13$  ms,  $T_e = 0.28$  ms,  $T_v = 0.04$  ms

one operation during authentication phase such as  $SK_{ac} = H(\hat{C}_i || N_a || k_a)$ . The computational costs for corresponding operations are  $T_m = (3 * 0.43) \approx 0.13$  ms,  $T_h = (3 * 0.01) \approx 0.03$  ms,  $T_e = (4 * 0.07) \approx 0.28$  ms, and  $T_v = (4 * 0.01) \approx 0.04$  ms. Hence, the total computational cost of the proposed protocol is  $T_c = T_m + T_h + T_e + T_v$ , which is  $T_c = (0.09 + 0.39 + 1.12 + 0.04) \approx 1.64$  ms. The block creation time depends on the frequency of data requests from the consumers so block verification time is considerably low. Table 5 clearly describes the operations involved during the registration and authentication phases of our proposed protocol. It is important to mention here that the concatenation operations are omitted due to negligible effect on computation as compared to other operations. From Fig. 5 it is illustrated that our proposed protocol is computationally efficient by comparing it with [26], [31], and [34] having the time cost  $\approx 1.64$  ms. In comparison with other protocols, percentage improvement can be described by calculating the difference in times of protocols. For [26] our protocol has 46.24%, with [31] it has 83.23%, and with [34] it has 65.23% less computational costs.

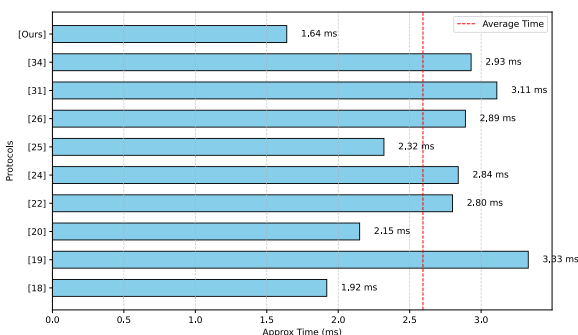


FIGURE 5. Computational cost comparison.

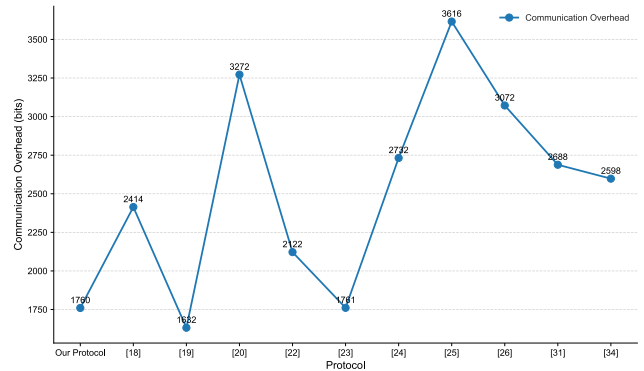


FIGURE 6. Communication cost comparison.

## 2) COMMUNICATION OVERHEAD ANALYSIS

We assumed the parameters to carry out secure communication among the system's participants to evaluate the communication cost for our proposed protocol. The cost of communication refers to the bits utilized to send messages among participants throughout the registration and authentication phases. As a result, communication overhead is obtained from the exchange of messages among participants. Certain assumptions are taken into account to help with the evaluation process. We assume that the consumer identity is 32 bits long, the timestamp is 8 bits long, the random number is 128 bits long, the elliptic curve point or hash function (SHA-256) is 256 bits long, and the ciphertext generated by the encryption/decryption function is 128 bits long. In our proposed protocol,  $C_i$  sends the two messages  $\mathcal{M}_c \leftarrow \{ID, R_E\}$  and  $\mathcal{M}_{c1} \leftarrow \{m_{c1}\}$  to  $B_A$ , which needs  $(32+128) + 160 = 320$  bits for message transmission. Similarly,  $C_i$  also sends two messages  $\mathcal{M}_{c2} \leftarrow \{R_c, m_c\}$  and  $\mathcal{M}_{c3} \leftarrow \{AT_{C_i}\}$  to  $\mathcal{P}_i$  which requires  $(128+256) + 256 = 640$  bits. In same way,  $B_A$  send messages to  $C_i$  such as  $\mathcal{M}_b \leftarrow \{\zeta_i, SK_{ac}\}$ ,  $\mathcal{M}_{b1} \leftarrow \{m_{a1}\}$ , and  $\mathcal{M}_{b2} \leftarrow \{AT_{C_i}\}$ , which requires  $(160 + 160) + 160 + 160 = 640$  bits for message transmission.  $\mathcal{P}_i$  sends message to  $B_A$  such as  $\mathcal{M}_p \leftarrow \{m_p\}$  which needs 160 bits to transmit message. So, the accumulative communication overhead for the proposed protocol is  $320 + 640 + 640 + 160 = 1760$  bits. Communication overhead for other protocols is calculated using the same method. Fig. 6 shows that the communication overhead of [18], [19], [20], [21], [22], [23], [24], [25], [34], and [35]. The efficiency of the protocol proposed in [21] is best compared to our proposed

protocol. However, the protocol proposed in [19] can not protect the privacy of data communication. Our protocol provides better security and privacy features as a trade-off and has greater efficiency compared to other schemes. With the comparison of these protocols, our protocol has less communication overhead from [26] and [31], which is 42.66% and 34.52%, respectively. Similarly, with the comparison of blockchain-enabled protocols our protocol has less communication overhead from [19] by 56.32%, [23] by 23.11%, and [25] by 37.43%.

The time spent for various cryptographic processes and blockchain tasks during protocol implementation is depicted in Figure. 7. We ran 50 transactions to record the execution times of activities like encryption, decryption, HMAC verification, block construction, and block verification. The image depicts how the execution times differ between the processes and how much variance occurs as shown in Figure. 7. The execution timings for encryption and decryption are rather consistent, whereas the times for HMAC generation, HMAC verification, and blockchain verification vary substantially. The blockchain creation process has the longest average execution time and the most substantial variability of any operation.

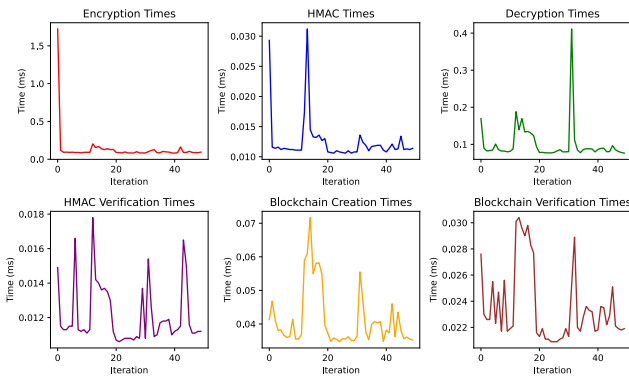


FIGURE 7. Time comparison of overall system operations.

**E. EFFICIENCY OF PROFIT DISTRIBUTION MECHANISMS**

We show the outcomes of the simulation profit distribution framework. The simulations were carried out in Python, with the libraries NumPy v1.23.5, Matplotlib v3.3.2, and Pandas v1.1.3. We calculated the Shapley value for data sources using the SHAP library. The dataset used is (<https://archive.ics.uci.edu/ml/machine-learning-databases/adult/>) from the UCI machine-learning library.

Figure. 8 shows how each provider influences data value for each of the five attributes. For attribute Age, Providers A through E’s contributions are high as compared to other attributes. Provider A contributes the most to Attribute Age, followed by B, C, D, and E. The other attributes follow the same trend, however, each data provider contributes differently. For example, Provider A still contributes the most to attribute Relationship, but Providers B and C’s contributions

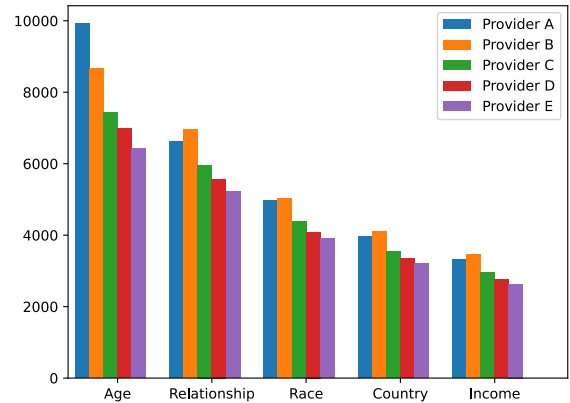


FIGURE 8. Value of Data according to attributes provided by data providers.

are considerably closer to Provider A’s contribution than they were for attribute Age. The contribution of each data provider depends on the data value they hold with different attributes.

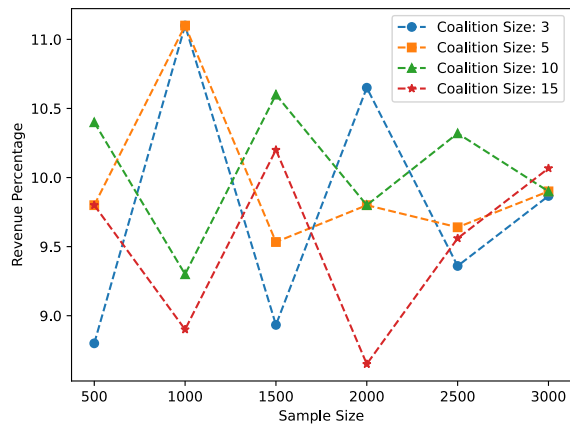


FIGURE 9. Effect of the coalition and sample size on revenue percentage.

Revenue percentages vary on coalition size as shown in Figure. 9, which indicates that increasing sample size improves revenue % for all coalition sizes. A larger sample size yields more accurate data, which is used to predict coalition revenue. It signifies that larger coalitions generate higher revenue percentages, while the variations are small compared to the generic pattern of increasing revenue % with sample size. The marginal contribution for each data provider is the difference between the total value generated by all players (coalition value) and the value generated by the coalition excluding the current player, multiplied by the monetary value. Cumulative contribution is all marginal contributions.

$$V(S) = \sum_{i \in S} w_i(x_i) - \sum_{i \notin S} w_i(\bar{x}_i) \tag{35}$$

The weight (importance) of attribute *i* in the decision model is represented by *w<sub>i</sub>* in Equation 35. The attribute weights are considered to add up to a total of one. The value of attribute *i* for data provider *j* is represented as *x<sub>ij</sub>*. *x<sub>age,j</sub>*, e.g., reflects the age of data provider *j*, and *x<sub>race,j</sub>*, race of data provider *j*.

The difference between the coalition value with and without data provider  $i$  is defined as the marginal contribution.

$$MC_i = (\phi_i(v(S_i \cup i)) - \phi_i(v(S_i))) \cdot M \quad (36)$$

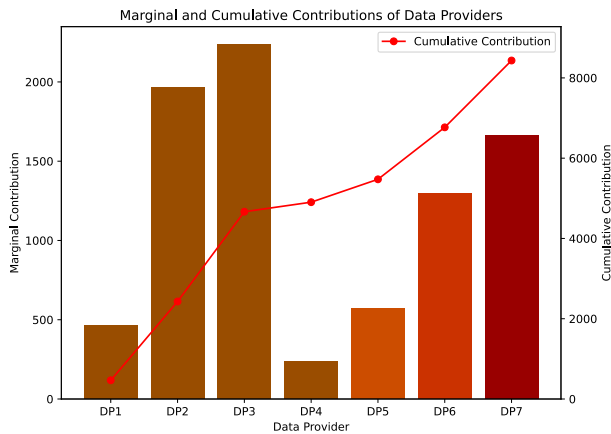


FIGURE 10. Marginal contribution of data providers to the coalition.

The sum of all prior marginal contributions is the cumulative contribution. Figure. 10 depicts the marginal contributions of each data provider, with DP2 and DP3 having the highest monetary value contributions.

$$CC_i = \sum_{j=1}^i MC_j \quad (37)$$

The data provider’s marginal contribution is  $MC_j$ , whereas the cumulative contribution is  $CC_i$ . Figure. 10 shows the  $MC_j$  and  $CC_i$  of seven data providers (DP1 to DP7) to a monetary value (100). In Figure. 10, DP2 and DP3 have the biggest  $MC_j$  which implies their revenues. The line shows the data sources’  $CC_i$  relatively increase for other data providers (DP4 to DP7) and measured as an overall contribution.

### F. VERIFICATION OF REVENUE SCHEME

We examine the relationship between third-party revenue ratio and model accuracy to assess data provider incentives. Our main goal is to emphasize the importance of incentive mechanisms and the relationship between data provider quality and revenue.

We used intelligent algorithms to construct a prediction model to study how shared data affects data provider revenue. We used the support vector machine (SVM), which is known for its data classification and real-world applicability.

The adult dataset used in our study came from the UCI Machine Learning Repository. To evaluate the prediction model, we chose 6,043 training data samples and 932 test data samples.

We divided the datasets into segments with 500-4500 data samples from data providers. We performed 100 simulations to calculate the data provider’s revenue distribution under different data sample sizes for each value. With more data samples, data provider P1 offered more data, therefore we

applied the method to calculate the distribution of revenue under different proportions, as shown in Figure 11.

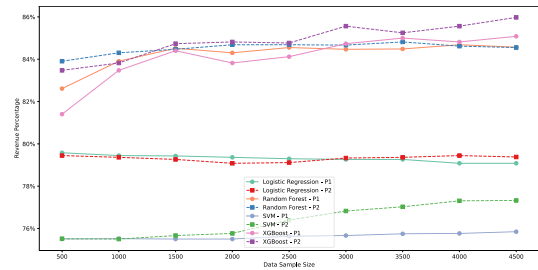


FIGURE 11. Data providers to the coalition.

According to the design goal and incentive scheme, increasing the quantity of data samples shared by data provider P1 increases revenue for both the data provider and others under unchanging conditions. When the data provider increases shared data samples, the learning algorithm’s prediction model improves. This shows that the data provider’s contribution increases, resulting in more revenue.

### G. PERFORMANCE EVALUATION OF SMART CONTRACT TRANSACTIONS

To test the usefulness of our approach, we built a blockchain network and executed our data-sharing mechanism in real-time. We utilized Remix IDE to create the solidity smart contract for this purpose. We have used different addresses to simulate the operations. Blockchain authenticators use \$0xd9145CCE52D...9943F39138\$ this address to register both data providers and consumers. The first data provider is registered with the address \$0xAb8483F64d9C...5835cb2\$ and the data consumer with the address \$0x5B38Da6a701c...6beddC4\$ as shown in Figure. 12. The cost associated with the

```

"from": "0xd9145CCE52D386f254917e481eB44e9943F39138",
"topic": "0x20f06cf92183fb3bd87d074ec788beb83367293faF2388c274b4bad9aeca725",
"event": "ConsumerRegistered",
"args": {
  "0": "0x5B38Da6a701c568545dCfC803FcB875f56beddC4",
  "consumerAddress": "0x5838Da6a701c568545dCfC803FcB875f56beddC4"
}

"from": "0xd9145CCE52D386f254917e481eB44e9943F39138",
"topic": "0x70abce74777b3838ae60a33a6b9a87d9d25532668fe4fea548554c55868579c0",
"event": "ProviderRegistered",
"args": {
  "0": "0xAb8483F64d9C6d1EcF9b849Ae677dD3315835cb2",
  "providerAddress": "0xAb8483F64d9C6d1EcF9b849Ae677dD3315835cb2"
}
    
```

FIGURE 12. Registration of data provider and consumer.

registrations of data providers and consumers is 23745 GAS. Here GAS refers to the execution cost of a transaction in a blockchain network. Figure. 13 illustrate GAS, transaction, and execution costs for the operations involved in the proposed system. It shows that uniform trend for smart contract transaction submission and execution cost during the registration of the participants. Moreover, it consumes

considerably less amount of GAS for withdrawal and data count as compared to GAS consumed for identity operations.

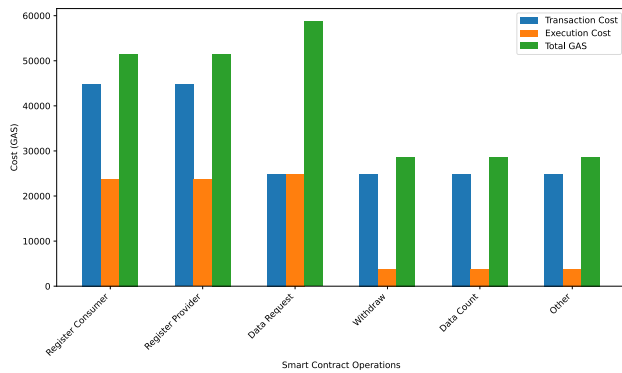


FIGURE 13. Transactions execution cost (GAS).

## H. DISCUSSION

The proposed protocol solves the problems faced in data sharing in IIoT against various security and privacy threats. By using ECC along with its properties and HMAC it enables the participants to share important data anonymously and securely. Using a consortium blockchain for participant approval ensures the safe and fair utilization of data in our system. We devised an incentive system for flexible benefit distribution across the platforms to encourage consortium participants to share trustworthy data and collaborate. However, the proposed framework can be extended to other systems such as in vehicular networks [51] for resource allocation as well as for providing efficient security protocol. Similarly, in IoT networks, the major challenge that can arise is the authentication of devices so this proposed protocol can be implemented in this scenario [52]. Apart from the applicability of the proposed framework in this paper, we plan to include a machine learning model to predict the incentive more efficiently through Shapley value [53]. Because IIoT generates an extensive amount of data with dynamic properties there is a margin of improvement for the incentive prediction and distribution.

## VIII. CONCLUSION AND FUTURE WORK

In this paper, we proposed the framework for secure data sharing and incentive distribution mechanism for the participants of the system. We designed the authentication and token generation protocol using elliptic curve properties and keyed HMAC that have considerably low computational and communication costs that show the efficiency and effectiveness of our protocol. Moreover, extensive simulations have been performed using AVISPA and Scyther simulation tools to verify the security features of the protocol. The simulation results show that the protocol is resilient against various adversarial attacks. We also design a profit generation and distribution among the participants of the data-sharing system. We have employed Shapley Value to distribute the profit among the participants of the system. The distribution

of profit is based on the data contribution of each data provider to the coalition. We have demonstrated that our profit distribution mechanism efficiently distributes the profit among the data providers with fairness.

In future work, we will apply our security framework to other areas of fields such as vehicular networks, and employ our incentive distribution mechanisms to the energy trading framework.

## REFERENCES

- [1] R. Huo, S. Zeng, Z. Wang, J. Shang, W. Chen, T. Huang, S. Wang, F. R. Yu, and Y. Liu, "A comprehensive survey on blockchain in industrial Internet of Things: Motivations, research progresses, and future challenges," *IEEE Commun. Surveys Tuts.*, vol. 24, no. 1, pp. 88–122, 1st Quart., 2022.
- [2] D. Hamouda, M. A. Ferrag, N. Benhamida, and H. Seridi, "PPSS: A privacy-preserving secure framework using blockchain-enabled federated deep learning for industrial IoTs," *Pervas. Mobile Comput.*, vol. 88, Jan. 2023, Art. no. 101738.
- [3] A. Hazra, P. Rana, M. Adhikari, and T. Amgoth, "Fog computing for next-generation Internet of Things: Fundamental, state-of-the-art and research challenges," *Comput. Sci. Rev.*, vol. 48, May 2023, Art. no. 100549.
- [4] G. K. Pandey, D. S. Gurjar, H. H. Nguyen, and S. Yadav, "Security threats and mitigation techniques in UAV communications: A comprehensive survey," *IEEE Access*, vol. 10, pp. 112858–112897, 2022.
- [5] S. Misra, C. Roy, T. Sauter, A. Mukherjee, and J. Maiti, "Industrial Internet of Things for safety management applications: A survey," *IEEE Access*, vol. 10, pp. 83415–83439, 2022.
- [6] B. Alotaibi, "A survey on industrial Internet of Things security: Requirements, attacks, AI-based solutions, and edge computing opportunities," *Sensors*, vol. 23, no. 17, p. 7470, Aug. 2023.
- [7] R. Kumar, P. Kumar, R. Tripathi, G. P. Gupta, A. K. M. N. Islam, and M. Shoruffzaman, "Permissioned blockchain and deep learning for secure and efficient data sharing in industrial healthcare systems," *IEEE Trans. Ind. Informat.*, vol. 18, no. 11, pp. 8065–8073, Nov. 2022.
- [8] J. A. Alzubi, O. A. Alzubi, A. Singh, and M. Ramachandran, "Cloud-IIoT-based electronic health record privacy-preserving by CNN and blockchain-enabled federated learning," *IEEE Trans. Ind. Informat.*, vol. 19, no. 1, pp. 1080–1087, Jan. 2023.
- [9] T. Li, J. Zhang, Y. Shen, and J. Ma, "Hierarchical and multi-group data sharing for cloud-assisted industrial Internet of Things," *IEEE Trans. Services Comput.*, 2023.
- [10] S. Halder and T. Newe, "Secure time series data sharing with fine-grained access control in cloud-enabled IIoT," in *Proc. IEEE/IFIP Netw. Operations Manage. Symp.*, Apr. 2022, pp. 1–9.
- [11] J. Sengupta, S. Ruj, and S. D. Bit, "FairShare: Blockchain enabled fair, accountable and secure data sharing for industrial IoT," *IEEE Trans. Netw. Service Manage.*, 2023.
- [12] R. Ma, L. Zhang, Q. Wu, Y. Mu, and F. Rezaeibagha, "BE-TRDSS: Blockchain-enabled secure and efficient traceable-revocable data-sharing scheme in industrial Internet of Things," *IEEE Trans. Ind. Informat.*, 2023.
- [13] F. Yang, Y. Qiao, M. Z. Abedin, and C. Huang, "Privacy-preserved credit data sharing integrating blockchain and federated learning for Industrial 4.0," *IEEE Trans. Ind. Informat.*, vol. 18, no. 12, pp. 8755–8764, Dec. 2022.
- [14] M. Rizwan, M. N. Sohail, A. Asheralieva, A. Anjum, and P. Angin, "SAID: ECC-based secure authentication and incentive distribution mechanism for blockchain-enabled data sharing system," in *Proc. IEEE Int. Conf. Blockchain*, Dec. 2021, pp. 530–537.
- [15] F. Zhang, H. Wang, L. Zhou, D. Xu, and L. Liu, "A blockchain-based security and trust mechanism for AI-enabled IIoT systems," *Future Gener. Comput. Syst.*, vol. 146, pp. 78–85, Sep. 2023.
- [16] A. Makkar, T. W. Kim, A. K. Singh, J. Kang, and J. H. Park, "SecureIIoT environment: Federated learning empowered approach for securing IIoT from data breach," *IEEE Trans. Ind. Informat.*, vol. 18, no. 9, pp. 6406–6414, Sep. 2022.
- [17] M. Adil, M. A. P. Mahmud, A. Z. Kouzani, and S. Khoo, "Energy trading among electric vehicles based on Stackelberg approaches: A review," *Sustain. Cities Soc.*, vol. 75, Dec. 2021, Art. no. 103199.

- [18] Y. Liu, A. Liu, Y. Xia, B. Hu, J. Liu, Q. Wu, and P. Tiwari, "A blockchain-based cross-domain authentication management system for IoT devices," *IEEE Trans. Netw. Sci. Eng.*, vol. 11, no. 1, pp. 115–127, 2024.
- [19] K. Wang, K. Sun, J. Dong, L. Sha, and F. Xiao, "AP-CDE: Cost-efficient authentication protocol for cross-domain data exchange in IIoT," *IEEE Syst. J.*, vol. 17, no. 3, pp. 3882–3893, Sep. 2023.
- [20] F. Wang, J. Cui, Q. Zhang, D. He, C. Gu, and H. Zhong, "Blockchain-based lightweight message authentication for edge-assisted cross-domain industrial Internet of Things," *IEEE Trans. Dependable Secure Comput.*, 2023.
- [21] A. A. Khan, S. Bourouis, M. M. Kamruzzaman, M. Hadjouni, Z. A. Shaikh, A. A. Laghari, H. Elmannai, and S. Dhabbi, "Data security in healthcare industrial Internet of Things with blockchain," *IEEE Sensors J.*, vol. 23, no. 20, pp. 25144–25151, 2023.
- [22] T. Li, H. Wang, D. He, and J. Yu, "Designated-verifier aggregate signature scheme with sensitive data privacy protection for permissioned blockchain-assisted IIoT," *IEEE Trans. Inf. Forensics Security*, vol. 18, pp. 4640–4651, 2023.
- [23] Y. Zhang, D. He, P. Vijayakumar, M. Luo, and X. Huang, "SAPFS: An efficient symmetric-key authentication key agreement scheme with perfect forward secrecy for industrial Internet of Things," *IEEE Internet Things J.*, vol. 10, no. 11, pp. 9716–9726, 2023.
- [24] K. Siva Sai, R. Bhat, M. Hegde, and J. Andrew, "A lightweight authentication framework for fault-tolerant distributed WSN," *IEEE Access*, vol. 11, pp. 83364–83376, 2023.
- [25] M. Hammad, A. Badshah, G. Abbas, H. Alasmary, M. Waqas, and W. A. Khan, "A provable secure and efficient authentication framework for smart manufacturing industry," *IEEE Access*, vol. 11, pp. 67626–67639, 2023.
- [26] M. A. Saleem, X. Li, M. F. Ayub, S. Shamshad, F. Wu, and H. Abbas, "An efficient and physically secure privacy-preserving key-agreement protocol for vehicular ad-hoc network," *IEEE Trans. Intell. Transp. Syst.*, 2023.
- [27] Y. Wang, T. Che, X. Zhao, T. Zhou, K. Zhang, and X. Hu, "A blockchain-based privacy information security sharing scheme in industrial Internet of Things," *Sensors*, vol. 22, no. 9, p. 3426, Apr. 2022.
- [28] J. Oh, J. Lee, M. Kim, Y. Park, K. Park, and S. Noh, "A secure data sharing based on key aggregate searchable encryption in fog-enabled IoT environment," *IEEE Trans. Netw. Sci. Eng.*, vol. 9, no. 6, pp. 4468–4481, Nov. 2022.
- [29] J. Lee, J. Oh, and Y. Park, "A secure and anonymous authentication protocol based on three-factor wireless medical sensor networks," *Electronics*, vol. 12, no. 6, p. 1368, Mar. 2023.
- [30] A. Jagruthi, K. Vikas, G. Nookaraju, K. R. Teja, G. Sanjana, and M. A. Jabbar, "Enhancing data spillage in multi-cloud storage services," in *Proc. 13th Int. Conf. Comput. Commun. Netw. Technol. (ICCCNT)*, Oct. 2022, pp. 1–4.
- [31] F. Yi, L. Zhang, L. Xu, S. Yang, Y. Lu, and D. Zhao, "WSNEAP: An efficient authentication protocol for IIoT-oriented wireless sensor networks," *Sensors*, vol. 22, no. 19, p. 7413, Sep. 2022.
- [32] C. Wang, S. Wang, X. Cheng, Y. He, K. Xiao, and S. Fan, "A privacy and efficiency-oriented data sharing mechanism for IIoTs," *IEEE Trans. Big Data*, vol. 9, no. 1, pp. 174–185, Feb. 2023.
- [33] X. Luo, H. Wang, J. Dong, C. Zhang, and T. Wu, "Achieving privacy-preserving data sharing for dual clouds," in *Proc. IEEE Int. Conf. Internet Things (iThings) IEEE Green Comput. Commun. (GreenCom) IEEE Cyber. Phys. Social Comput. (CPSCom) IEEE Smart Data (SmartData) IEEE Congr. Cybermatics (Cybermatics)*, Aug. 2022, pp. 139–146.
- [34] M. Tanveer, A. Alkhayyat, A. U. Khan, N. Kumar, and A. G. Alharbi, "REAP-IIoT: Resource-efficient authentication protocol for the industrial Internet of Things," *IEEE Internet Things J.*, vol. 9, no. 23, pp. 24453–24465, Dec. 2022.
- [35] W. Wang, Z. Han, M. Alazab, T. R. Gadekallu, X. Zhou, and C. Su, "Ultra super fast authentication protocol for electric vehicle charging using extended chaotic maps," *IEEE Trans. Ind. Appl.*, vol. 58, no. 5, pp. 5616–5623, Sep. 2022.
- [36] C. Zhang, T. Shen, and F. Bai, "Toward secure data sharing for the IoT devices with limited resources: A smart contract-based quality-driven incentive mechanism," *IEEE Internet Things J.*, 2022.
- [37] T. Mai, H. Yao, J. Xu, N. Zhang, Q. Liu, and S. Guo, "Automatic double-auction mechanism for federated learning service market in Internet of Things," *IEEE Trans. Netw. Sci. Eng.*, vol. 9, no. 5, pp. 3123–3135, Sep. 2022.
- [38] X. Chen, G. Zhu, H. Ding, L. Zhang, H. Zhang, and Y. Fang, "End-to-end service auction: A general double auction mechanism for edge computing services," *IEEE/ACM Trans. Netw.*, vol. 30, no. 6, pp. 2616–2629, Dec. 2022.
- [39] W. Borjigin, K. Ota, and M. Dong, "Multiple-Walrasian auction mechanism for tree valuation service in NFV market," *IEEE Trans. Computat. Social Syst.*, vol. 10, no. 1, pp. 61–71, Feb. 2023.
- [40] J. Zhang, J. Hei, and H. Tan, "Edge pricing mechanisms under cloud tiered pricing," in *Proc. 8th Int. Conf. Big Data Comput. Commun. (BigCom)*, Aug. 2022, pp. 54–62.
- [41] K. Zhu, L. Huang, J. Nie, Y. Zhang, Z. Xiong, H.-N. Dai, and J. Jin, "Privacy-aware double auction with time-dependent valuation for blockchain-based dynamic spectrum sharing in IIoT systems," *IEEE Internet Things J.*, vol. 10, no. 8, pp. 6756–6768, 2023.
- [42] H. Qiu, K. Zhu, N. C. Luong, C. Yi, D. Niyato, and D. I. Kim, "Applications of auction and mechanism design in edge computing: A survey," *IEEE Trans. Cognit. Commun. Netw.*, vol. 8, no. 2, pp. 1034–1058, Jun. 2022.
- [43] J. Kang, J. Wen, D. Ye, B. Lai, T. Wu, Z. Xiong, J. Nie, D. Niyato, Y. Zhang, and S. Xie, "Blockchain-empowered federated learning for healthcare metaverses: User-centric incentive mechanism with optimal data freshness," *IEEE Trans. Cognit. Commun. Netw.*, vol. 10, no. 1, pp. 348–362, Feb. 2024.
- [44] S. Li and S. Qu, "The three-level supply chain finance collaboration under blockchain: Income sharing with Shapley value cooperative game," *Sustainability*, vol. 15, no. 6, p. 5367, Mar. 2023.
- [45] X. Yang, S. Xiang, C. Peng, W. Tan, Z. Li, N. Wu, and Y. Zhou, "Federated learning incentive mechanism design via Shapley value and Pareto optimality," *Axioms*, vol. 12, no. 7, p. 636, Jun. 2023.
- [46] T. K. Dang, P. T. Tran-Truong, and N. T. H. Trang, "An enhanced incentive mechanism for crowdsourced federated learning based on contract theory and Shapley value," in *Future Data and Security Engineering. Big Data, Security and Privacy, Smart City and Industry 4.0 Applications*, T. K. Dang, J. Küng, and T. M. Chung, Eds. Singapore: Springer, 2023, pp. 18–33.
- [47] Y. Chai and X.-J. Zeng, "Shapley value-based computation offloading for edge computing," *IEEE Trans. Veh. Technol.*, vol. 72, no. 7, pp. 9448–9458, 2023.
- [48] D. Dolev and A. Yao, "On the security of public key protocols," *IEEE Trans. Inf. Theory*, vol. IT-29, no. 2, pp. 198–208, Mar. 1983.
- [49] R. Canetti and H. Krawczyk, "Universally composable notions of key exchange and secure channels," in *Advances in Cryptology—EUROCRYPT 2002*, L. R. Knudsen, Ed. Berlin, Germany: Springer, 2002, pp. 337–351.
- [50] M. Abdalla, P.-A. Fouque, and D. Pointcheval, "Password-based authenticated key exchange in the three-party setting," in *Public Key Cryptography—PKC 2005*, S. Vaudenay, Ed. Berlin, Germany: Springer, 2005, pp. 65–84.
- [51] A. Ribeiro, J. B. D. da Costa, G. P. R. Filho, L. A. Villas, D. L. Guidoni, S. Sampaio, and R. I. Meneguette, "HARMONIC: Shapley values in market games for resource allocation in vehicular clouds," *Ad Hoc Netw.*, vol. 149, Oct. 2023, Art. no. 103224.
- [52] R. Kumar, D. Javeed, A. Aljuhani, A. Jolfaei, P. Kumar, and A. K. M. N. Islam, "Blockchain-based authentication and explainable AI for securing consumer IoT applications," *IEEE Trans. Consum. Electron.*, 2024.
- [53] M. Shafiq, R. Yadav, A. R. Javed, and S. A. H. Mohsin, "CoopGBFS: A federated learning and game-theoretic based approach for personalized security, recommendation in 5G beyond IIoT environments for consumer electronics," *IEEE Trans. Consum. Electron.*, 2024.



**MUHAMMAD NOMAN SOHAIL** received the bachelor's degree in information technology from the University of Sargodha, in 2018, and the master's degree in information security from COMSATS University Islamabad, Pakistan, in 2021. He is currently a Lecturer with the Department of Computer Science, The University of Lahore, Sargodha Campus. Previously, he was a Visiting Lecturer with the University of Sargodha. His research interests include information security, blockchain, the Internet of Vehicles, data privacy, and cryptography.





**ADEEL ANJUM** received the Ph.D. degree in computer science from Polytech Nantes, Nantes, France, in 2013. He is currently a Professor and the Director of the Institute of Information Technology, Quaid-i-Azam University, Islamabad, Pakistan. He has several publications and authored a book on data privacy. His research interest includes AI-based data privacy. He served on the technical program committees for various international conferences.



**IFTIKHAR AHMED SAEED** received the degree in information technology from the College of Information and Electrical Engineering, China Agricultural University, and the master's degree in computer science and engineering technology in Pakistan. He is currently an Assistant Professor with the Department of Computer Science, The University of Lahore. His research interests include the design and development of soil sensors, multi-sensor techniques, multi-sensor fusion, and the agricultural Internet of Things (IoT).



**MADIHA HAIDER SYED** received the Ph.D. degree in computer science from Florida Atlantic University, USA, in 2019. She is currently an Assistant Professor with the Institute of Information Technology, Quaid-i-Azam University, Pakistan. Her research interests include cloud computing, security, privacy, software architecture, the IoT, machine learning, and deep learning. She was a recipient of the prestigious Fulbright Scholarship, in 2014, for the Ph.D. degree.



**AXEL JANTSCH** (Senior Member, IEEE) received the Dipl.-Ing. and Ph.D. degrees in computer science from TU Wien, Vienna, Austria, in 1987 and 1992, respectively. From 1997 to 2002, he was an Associate Professor with the KTH Royal Institute of Technology, Stockholm. From 2002 to 2014, he was a Full Professor of electronic systems design with KTH. Since 2014, he has been a Professor of systems on chips with the Institute of Computer Technology, TU Wien. He has published five books as an editor and one as the author and has more than 300 peer-reviewed contributions in journals, books, and conference proceedings. He has given more than 100 invited presentations at conferences, universities, and companies. His current research interests include systems on chips, self-aware cyber-physical systems, and embedded machine learning.



**SEMEEN REHMAN** (Member, IEEE) received the Habilitation degree in embedded systems from the Faculty of Electrical Engineering and Information Technology, Technische Universität Wien (TU Wien), in October 2020, and the Ph.D. degree from KIT, Germany. She is currently with TU Wien as an Assistant Professor. She has coauthored one book, multiple book chapters, and more than 70 publications in premier journals and conferences. Her main research interests include dependable and energy-efficient embedded systems, approximate computing, security, and CPS/IoT. She received the CODES+ISSS 2011 and the 2015 Best Paper Awards, the 2017 DATE Best Paper Award Nomination, the HiPEAC Paper Awards, the DAC Richard Newton Young Student Fellow Award, and the Research Student Award from KIT. She served as the topic track chair/the co-chair and has served as the TPC for multiple premier conferences on design automation and embedded systems.

...