Abstract—The rapid growth of the Internet of Things (IoT) has brought big data analytics (BDA) to the forefront of IoT research. However, existing efforts have primarily focused on enhancing data mining efficiency, often neglecting the preferences and gains of stakeholders involved in BDA. To address this gap, this paper introduces a novel approach that combines game theory and auction mechanisms to incentivize participation and optimize benefits for all stakeholders in the context of IoT's big data analytics (BDA-IoT). Our proposed model encourages voluntary engagement and employs an auction mechanism to establish a structured incentive framework. Through simulation and analysis, we demonstrate the viability of our model in fostering stakeholder cooperation and maximizing benefits, underscoring its practicality and rationality.

I. INTRODUCTION

The emergence of the Internet of Things (IoT), first coined by Kevin Ashton in 1999 [1],[2],[3], has revolutionized the concept of networked devices. IoT has now become an integral part of our lives, with over 14.4 billion IoT devices connected to the Internet by 2022 [4],[5].This advanced worldwide network infrastructure possesses autonomous configuration abilities. It serves as a link between the tangible realm of Operations Technology (OT) and devices, as well as the virtual realm of Information Technology (IT) and applications [6]. These devices generate substantial volumes of valuable data, and the utilization of big data analytics in IoT (BDA-IoT) holds the potential to significantly enhance the value of IoT systems by enabling predictive insights. However, the decentralized nature of IoT devices and the sheer magnitude of data present challenges in their effective management and analysis. To address this, cloud storage technology has been proposed for IoT data storage, allowing for unified data management and subsequent analysis[7]. Public cloud storage providers (CSPs) have been suggested for IoT data storage to enhance data integration, accessibility, and analysis[8],[9]. Although preliminary research highlights the viability and necessity of BDA-IoT, existing methodologies often assume universal stakeholder cooperation, failing to consider individual motivations and incentives.

A. CHALLENGES AND TECHNIQUES

Challenges: This paper addresses several key challenges in the practical application of BDA-IoT:

Stakeholder Motivation: Existing research has primarily concentrated on optimizing technical aspects of data analytics, often neglecting the motivations and incentives of stakeholders. This creates a significant gap between theoretical solutions and their real-world implementation, as the willingness of stakeholders to participate plays a pivotal role.

Lack of Incentives: While there is acknowledgment of the importance of incentivizing participation, concrete methods for providing incentives to stakeholders, especially data owners, are lacking. Without compelling incentives, stakeholders may be hesitant to contribute their resources to BDA-IoT.

Techniques: To address the above challenges, we propose the following techniques:

Game Theory Framework: We adopt a game theory approach to model the interactions between stakeholders in the BDA-IoT ecosystem. Game theory provides a formalized structure to analyze the strategic decisions made by stakeholders while considering their rationality and self-interest.

Auction Mechanism for Incentives: To motivate stakeholder participation, we introduce an auction mechanism. In this model, CSPs act as auctioneers offering data owners participation incentives in the form of storage discounts. By allowing CSPs to compete for data owners' engagement, this mechanism stimulates the willingness of data owners to contribute their data.

Collaborative Benefits: The auction mechanism creates a collaborative environment where CSPs, data owners, and the Analytic Party (AN) share the benefits. This collaborative approach encourages stakeholder engagement as they can all reap the rewards of successful BDA-IoT implementation.

Incentive Structure Design: By structuring the auction in a way that CSPs offer storage discounts and data owners bid for participation, we establish a balanced incentive structure that aligns the interests of all stakeholders.

B. AUCTION MECHANISM FOR PARTICIPATION

Our proposed methodology leverages game theory and the auction mechanism to induce voluntary participation among stakeholders. Specifically, CSPs act as auctioneers, offering data owners attractive storage discounts as participation incentives. Data owners, motivated by these incentives, participate in the auction by bidding for the opportunity to engage in BDA-IoT. The second-highest bidder (data owner) is selected to participate, ensuring their privacy while still enabling analysis. This auction mechanism not only stimulates stakeholder involvement but also creates a fair and competitive environment that maximizes the benefits for all parties involved.

II. RELATED WORK

A. Big Data Analytics in IoT

The integration of big data analytics (BDA) and the Internet of Things (IoT) has the potential to improve the value of IoT systems. Marjani et al. proposed a framework for BDA-IoT that combines IoT and data analytics to provide an effective model. Sun et al. developed a smart grid-oriented big data analytic scheme that can effectively forecast power consumption through unique pricing mechanisms. Lv et al. proposed a 6G-based big data analytic system that can significantly increase the access rate of large-scale IoT devices and provide the underlying transmission guarantee for BDA-IoT. These schemes demonstrate that BDA-IoT can have a catalytic effect on IoT development.

However, the successful implementation of these schemes relies on the participation of stakeholders. Abu-Elkheir et al. constructed a federated data-centric IoT data management framework that links IoT data and potential IoT applications to improve data availability. Shafagh et al. devised a blockchain-based scheme to achieve distributed access control and data management, which enables secure data sharing and IoT data management. Ayoade et al. proposed a decentralized data management system for IoT devices that uses smart contracts and audit trails to enforce data access. This scheme allows multiple parties to specify rules to govern their interactions independently, thus improving data transparency and availability.

While there is a growing body of work on the framework and practical significance of BDA-IoT, few studies have investigated the willingness of stakeholders to participate in BDA-IoT. This is a critical factor that determines whether the system can be deployed smoothly. Therefore, it is important to investigate the interaction behavior between system participants.

System Model and Problem Formulation

Given a set of Big Data Center(s) (DCs) depicted as D = {D_1,D_2,...,D_k},where each D_i is a bundle consisting of some CSP centers. There are two possible cases for any pair of (D_i,D_j),depending on the number of CSPs in each D_i:

1. ∣ D_i ∣ = ∣ D_j ∣
2. ∣ D_i ∣ = ∣ D_j ∣

We can generalize the representation as follows: D_k denotes a set of CSP centers, which can be expressed as D_k = {nD_1^i, nD_2^i, ..., $\mathrm{nD}_{k_i}^i$}, where k_i ∈ {1, 2, ...,y}, and i ∈ {1, 2, ...,z}. This concept is illustrated in Figure 1. Each CSP center, denoted as nD_j^i, is a part of D_i and holds a specific dataset. Users can request Data Analytic services from these CSP centers as needed. To encourage active participation in this model, it is assumed that when a CSP center, nD_j^i, becomes a part of Data Center D_i, the Analytic Party (AN) compensates it with a fee for providing data for analytical purposes.

The CSP center asks data owners if they agree to share their data for data analytics and participate in the BDA-IoT Auction. If yes, the CSP center gives the data owners a special discount on storage. Otherwise, the data owners are charged the original price for storage. The CSP center then classifies and manages the stored data. An Analytic Party (AN) sends a request to the CSP center for access to the data. The CSP center sends the authorized data to the AN for data analytics. The CSP center charges the AN a certain amount for the data. The AN conducts data analytics on the IoT data to explore its potential value.

If a data owner provides their data to Di for storage purposes only, nD_j^i will charge a fixed fee. However, if the data owner provides their data to Di for analytics purposes, an auction will be held with the other data owners (Oj's).
The payment of nD_j^i can be expressed as $p_j^i = a_j^i * \theta_j^i + \bar{a}_j^i * \theta_j^{-i}$, where a_j^i is an indicator function defined as:

$$a_j^i = \left\{\frac{1, \mathit{if\ data\ is\ given\ for\ analytics}}{0, \quad \mathit{Otherwise}}\right\}$$

And

$$a_j^{-i} = \left\{\frac{1, \mathit{if\ data\ is\ not\ given\ for\ analytics}}{0, \quad \mathit{Otherwise}}\right\}$$

A data center nD_j^i can earn money when a sealed-bid auction is run for data. The CSP center cloud charges a fixed amount, θ_j^i, and also makes a payment, θ_j^{-i}, when the data is found outside of the CSP center. Each CSP center in D has a private valuation, g_i , known only to them.

A CSP center may earn some money also when an auction is run. If winner of the Data Owner Auction(O_i) selects Data Center D_i, then the money $nD_j^i \in D_i$ earns is: $p_j^{-i} = b_j^i(\frac{1}{2}q^i) + b_j^{-i}(\frac{1}{2}q^i w_j^i) + \alpha$. Here, like the previous case, a_j^i and $\bar{a}_j^i$ are the indicator functions defined as:-

$$b_j^i = \left\{\frac{1, \mathit{if}\ nD_j^i\ \mathit{is\ the\ contributor\ in}\ D_i}{0, \quad \mathit{Otherwise}}\right\}$$

$$b_j^{-i} = \begin{Bmatrix} \dfrac{1, if\ nD_j^i\ is\ not\ the\ contributor\ in\ D_i}{0, \quad Otherwise} \end{Bmatrix}$$

In a BDA– IoT environment, a contributor nD_j^i is given a share of the amount won, which is called the incentive. $\frac{1}{2}q^i$ is given to the contributor directly, and the other $\frac{1}{2}q^i$ is divided among the other contributors in the Data Center based on their weights. The weight of a contributor is a measure of how often their data is invoked by other CSP centers. So, contributors with higher weights receive a larger share of the incentive. And α is the fee that Analytic Party(AN) pays to the CSP Center.

When a new nD_j^i joins the Data Center Di, it is assigned a small weight w_j^i. The weight increases by 1 if its data is invoked by other CSP Centers. The remaining amount of the incentive is divided by a proportional share mechanism. The total payment made by an nD_j^i over all transactions can be defined as:

$$p_j^{-i} = \sum_{i=1}^{\hat{k}_1} \quad p_j^i \; - \; \sum_{i=1}^{\hat{k}_2} \quad p_j^i$$

where kˆ1 is the number of transactions for which nD_j^i is charged some discount money and kˆ2 is the number of transactions where it earned some money.

Proposed Mechanism

In this section, we propose an auction based AaaS algorithm for the framework discussed in the earlier sections. The algorithm is called Auction Based Aaas (ABAaaS). There are five main components of the ABAaas algorithm:

1. `Main_Routine`
2. `Identify`
3. `Run_Auction`
4. `Set_Buyer_Price`
5. `Set_Seller_Price`

Algorithm 1 Main_Routine

for $t = 1, 2, . . .$ **do**
    Query is made for $nD^i \in D_k$
    if data is given for analytics to D_k **then**
/* Update the payment of nD^i */
    $nD^i \cdot p \leftarrow nD^i \cdot p + DC$
/* DC is some discount given to Data Owner */
    else
/* Search for Data Owners */
    $i = i$

    $S \leftarrow$ Identify $(O \setminus O_i)$
    $S', v', nD, i' \leftarrow$ Run_Auction (S)
    /* $S' \rightarrow$ which data owner will provide the data for analytics
    $v' \rightarrow$ Second price
    $nD \rightarrow$ which data center inside S' contributed
    $i' \rightarrow$ The index of S' */
    Set_Buyer_Price (O_i, v')
    Set_Seller_Price (S', v', nD, i')
    end if
end for
end

If the data is provided for analytics within the data center where the CSP center belongs, the CSP center gives a discount to the data owner. Otherwise, an auction is run. The Identify routine first searches for the data owners who can provide the data, and then the Run_Auction routine runs an auction based on the Vickrey auction framework (Vickrey, 1961) that is cast into a reverse auction setting in our problem. The Set_Buyer_Price routine sets the payment of the buyers, and the Set_Seller_Price routine sets the payment of the CSP who contributes to the Big Data Analytics and the payment of all other CSP centers in the set Dk.

Algorithm 2 Identify (S^-)

$S \leftarrow \varphi$
for $i = 1$ *to* $|S^-|$ **do**
    if data available to S^-_i **then**
    $S \leftarrow S \cup \{S^-_i\}$
    end if
end for
return S

Algorithm 3 Run_Auction (S)

$v \leftarrow \varphi$
/* Collection of Bids */

2: **for** i = 1 to $|S|$ **do**

3: $v_i \leftarrow bid(S_i)$

4: $v \leftarrow v \cup v_i$

5: **end for**

/* Selecting Winner */

6: $i = \text{argmin}_{j: vj \in v}\ v_j$

7: $i' = i$ *Remember the index of the CSP Center selected*

8: $S' \leftarrow S_i$

9: $S = S - S_i$

/* Selecting the second highest bidder for payment */

10: **for** i = 1 to $|S|$ **do**

11: $vi \leftarrow bid(Si)$ /* storing all bids except the winner in v_i */

12: $v \leftarrow v \cup \{v_i\}$

13: **end for**

/* Selecting the second highest bidder */

14: $i = \text{argmin}_{j: vj \in v}\ v_j$

15: $v' = v_i$

16: nD $\leftarrow$ *extract* (S') *Which CSP Center served the Big Data Analytics*

17: **return** (S', v', nD, i')

Algorithm 4

Set_Buyer_Price (nD_j^i, v')

1: $nD_j^i \cdot p = nD_j^i \cdot p + v'$

Algorithm 5 Set_Seller_Price (S', v', nD, i')

1: for j = 1 to $|S'|$ do

2: if $nD = nD_j$ then

3: $nD_j^{i'} \cdot p^- = nD_j^{i'} \cdot p^- + \frac{1}{2}v'$

/* p^- is the payment earned and p is the payment spend */

4: $nD_j^{i'} \cdot p^- + \frac{1}{2}v' \cdot \frac{1}{|S'|-1}$ (equal share)

5: or $nD_j^{i'} \cdot p^- = nD_j^{i'} \cdot p^- + \frac{v'}{2} \cdot \frac{nD_j^{i'} . w_j^{i'}}{\sum_{nD_j^{i'} \in S' - nD} nD_j^{i'} . w_j^{i'}}$ (proportional share)

6: end if

7: end for