A Dynamic Service Trading in a DLT-Assisted Industrial IoT Marketplace

Jiejun Hu, Martin Reed, *Member, IEEE*, Nikolaos Thomos, *Senior Member, IEEE*, Mays F. Al-Naday, *Member, IEEE*, and Kun Yang, *Senior Member, IEEE*

***Abstract*—With the increasing demand for digitalization and participation in Industry 4.0, new challenges have emerged concerning the market of digital services to compensate for the lack of processing, computation, and other resources within Industrial Internet of Things (IIoTs). At the same time, the complexity of interplay among stakeholders has grown in size, granularity, and variation of trust. In this paper, we consider an IIoT resource market with heterogeneous buyers such as manufacturer owners. The buyers interact with the resource supplier dynamically with specific resource demands. This work introduces a broker between the supplier and the buyers, equipped with Distributed Ledger Technologies (DLT) providing a service for market security and trustworthiness. We first model the DLT-assisted IIoT market analytically to determine an offline solution and understand the selfish interactions among different entities (buyers, supplier, broker). Considering the non-cooperative heterogeneous buyers in the dynamic market, we then follow an independent learners framework to determine an online solution. In particular, the decision-making procedures of buyers are modeled as a Partially Observable Markov Decision Process which is solved using independent Q-learning. We evaluate both the offline and online solutions with analytical simulations, and the results show that the proposed approaches successfully maximize players' satisfaction. The results further demonstrate that independent Q-learners achieve equilibrium in a dynamic market even without the availability of complete information and communication, and reach a better solution compared to that of centralized Q-learning.**

***Index Terms*—DLT, Dynamic Pricing, IIoT marketplace, Independent Learning, POMDP**

I. Introduction

The Industry 4.0 revolution has been widely accepted over the last decade. With the development of sensors, machine learning algorithms, and network technologies, the number of companies that are interested in digital transformation has grown drastically[1]. Industrial Internet of Things (IIoTs), as the foundation of Industrial 4.0, aids the digitalization and advances smarter manufacturing methods. However, IIoT technology alone does not provide all the necessary systems to enable the full Industry 4.0 that also relies upon business innovation. Indeed, work by market economists [1] reveals a complex intertwined nature of platform ecosystems that form the overall Industry 4.0. These platforms include: end users, such as automotive manufacturers or property management systems; infrastructure providers for communications, networks, cloud and edge computing; data prediction and monitoring systems; digital financial services; and, *brokers and agents*. This paper focuses on new technology that provides *trustworthy* brokers that provide the bridge between the infrastructure providers and end users through digital financial services.

An IIoT *end service provider*, for example AWS IoT [2], presents easy-to-use smart manufacturing applications to a factory owner, however, the end service providers or the factory perhaps do not own all the computation and connectivity resources to facilitate the deployment of IIoT applications. Consequently, *infrastructure providers* supply resources such as Internet connectivity and computational resources (e.g. Siemens Industrial Edge [3]) for processes like data analytics and storage. Thus, the end service provider composes an IIoT service by purchasing a set of resources from the infrastructure provider(s), which enables faster deployment and scaling of IIoT applications. As the technology develops, a new business model that allows inter-operation among different stakeholders is required. Thus, an IIoT ecosystem needs a marketplace, to automatically match the resource requirements of the end services with the resources of the infrastructure providers. The recent work by [1] show that brokers that mediate this marketplace between the entities are an important part of the whole ecosystem, as is common in most business sectors that rarely sell direct to the end-user. However, this marketplace has not yet been automated or standardised across the entities and the central contribution of this paper is to provide a secure and fair solution.

The infrastructure providers and the IIoT end services require a *secure* marketplace environment to trade with each other without revealing crucial business intelligence, such as transaction details, resource requirements, and especially the final price. This provides a *level playing field* as is one of the founding principles for an online platform for delivering a digital market as proposed by the EU [4]. A central part of this marketplace is that infrastructure providers and the IIoT end services need to *fairly* build a contract with each other, which means that either side of the marketplace has no prejudice (e.g., no matter if they are startups or monopoly) but only focuses on the nature of the marketplace (i.e., building a contract upon the relationship of supply and demand). After a contract is formed, the transactions require a *fast and automatic* method to be verified, which enables real-time deployment of IIoT end services. Last but not least, the entities in the marketplace desire *stable* trading and *trustworthy* relationships with one another. While an automated marketplace does not currently

J. Hu is with the Lancaster University Leipzig, Germany. E-mail: j.hu14@lancaster.ac.uk.

M. Reed, N. Thomos, M. Al-Naday, and K. Yang are with the school of Computer Science and Electronic Engineering, University of Essex, UK.

[1]https://www.bitkom.org/sites/default/files/2021-04/bitkom-charts-industrie-4.0-07-04-2021_final.pdf

exist for IIoT, Riasanow *et al.* show how brokers are an integral part of business ecosystems and give examples of brokers that exist in the Financial Sector [1] that would form a technology solution for IIoT if a system such as proposed in this paper were to exist.

Thus, the central contribution of this paper addresses a scenario where there is an infrastructure provider that supplies different types of resource combinations, i.e., a combination of the computing resource and network management to satisfy different resource requirements from different end services. We consider there are a number of end service providers (buyers) that provide a range of services, e.g., artificial intelligence, digital twining, industrial robotics, etc. Due to the sensitive data possibly shared amongst a number of the stakeholders, security issues in Industry 4.0 draw our attention. In this paper, we propose distributed ledger technology (DLT) [5] as a Service (DLTaaS) as a broker to promote the resource demand by providing a secure marketplace for the transactions between the buyers and the supplier. DLT is the generic term of a distributed database managed by multiple parties. Blockchain [6] is a type of DLT where the transactions are recorded in a block and each block is immutably interconnected by a hash function. DLT is one of the revolutionary technologies in Industry 4.0, which can improve the security, transparency, and privacy during data exchange and resource trading. The security of the system relies upon the inherent security within the DLT, but also requires security support, for example by deploying an IoT security solution such as SerIoT [7]. We adopt *Hashgraph* [8] in this paper as the main DLT instance.

Hashgraph is a new data structure, compared to Blockchain, and is based on a Directed Acyclic Graph (DAG). Hashgraph is particularly appealing because it is designed for fast micro transaction processing in IoT applications. Hashgraph reduces the resource requirement of the consensus mechanism by using an asynchronous Byzantine consensus mechanism instead of Proof-of-Work and Practical Byzantine Fault Tolerance (PBFT) [9]. The proposed scenario forms a typical supply chain within the economic sphere. To simplify the terminology, hereafter we will use the terms *supplier*, *broker*, and *buyers* to refer to the *infrastructure provider*, *DLT service*, and the *end service providers* in Fig. 1, respectively.

The entities in the supply chain are all self-interested, which means they all want to maximize their profit. The supplier and the broker exert *marketing efforts* to promote the sales that results in higher profit for the infrastructure providers. Meanwhile end-users benefit from increased consumer choice leading to competitive pricing advantages [4]. The broker benefits by either taking a small commission for sales, or through advertising in larger direct to consumer markets. The marketing effort of a supplier could be the advertising and reaction speed; whereas, the marketing effort of a broker could be the number of CPU cycles in operating the consensus mechanism. However, the marketing efforts of the supplier and the broker are *hidden* from each other, which leads to information asymmetry. Typically, contract theory [10] is adopted to solve such a problem. When the buyers joins the marketplace, they may estimate the resource demand according to the reputation of the resource/service (obtained by an online review and annual report). Since the supplier and the broker are only capable of generating limited resources, the buyers would propose the optimal price according to their resource demand to successfully obtain the resource. This leads to a competitive environment among the buyers. They would hide their optimal unit price from other buyers to maximize their profit. This renders centralized learning approaches inappropriate for such marketplaces.

Based on the information asymmetries, i.e., between the supplier and the broker, and among the buyers, we first study the problem considering a static scenario which allows for deploying the offline analytical approach presented in Section V. This helps us understand the interactions between the different entities. Then, in Section VI, we estimate the reputation and cast an online approach based on Partially Observable Markov Processes (POMDP). We adopt the independent learning agents and use the light-weight Q-learning algorithm to determine the optimal policy of the buyers, i.e., the optimal price of the services in the dynamic IIoT marketplace. We did not adopt a centralized Q-learning framework or approaches in between centralized and independent Q-learning [11], [12] because both have high computational complexity. Further, the competitive nature of the studied problem, where buyers hide their optimal price from the others, means that the above approaches cannot be used. The main contributions of this paper are summarized as follows:

1) We propose and model a DLT assisted IIoT marketplace. We investigate the DLTaaS analytically and define the marketing effort of DLT as the security effort. The dynamic service trading market enables fast deployment of IIoT amongst a broader set of stakeholders, which additionally aids digitalization and smart manufacturing solutions to factories without fundamental computation and communication resources.
2) We analyze the interactions between different stakeholders in the presence of information asymmetry, i.e., the supplier and the broker hide their marketing efforts, and the buyers do not share the marketing demand and the unit price with other buyers.
3) We model the supply chain as a POMDP and adopt an independent Q-learning approach where agents represent the buyers. A POMDP makes it possible to determine efficient buyers' policies in the challenging dynamic IoT marketplace and hence helps all entities in the marketplace to reach their maximum satisfaction without requiring communication or coordination overheads. Q-learning assisted trading automation empowers a healthy degree of rivalry between the buyers and the supplier.

In the following, we first review related works in Section II. Then, in Section III, we introduce the architecture of the marketplace and the consensus mechanism of Hashgraph. Next, we provide the considered system model in Section IV. We, then, propose the analytical model in Section V; following with the online approach of the problem through Q-learning in VI and VI-B. Our solution is evaluated extensively in Section VII to get an understanding of the influence of the various system parameters to system's performance. Finally, we draw

This article has been accepted for publication in IEEE Transactions on Network and Service Management. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TNSM.2022.3191979

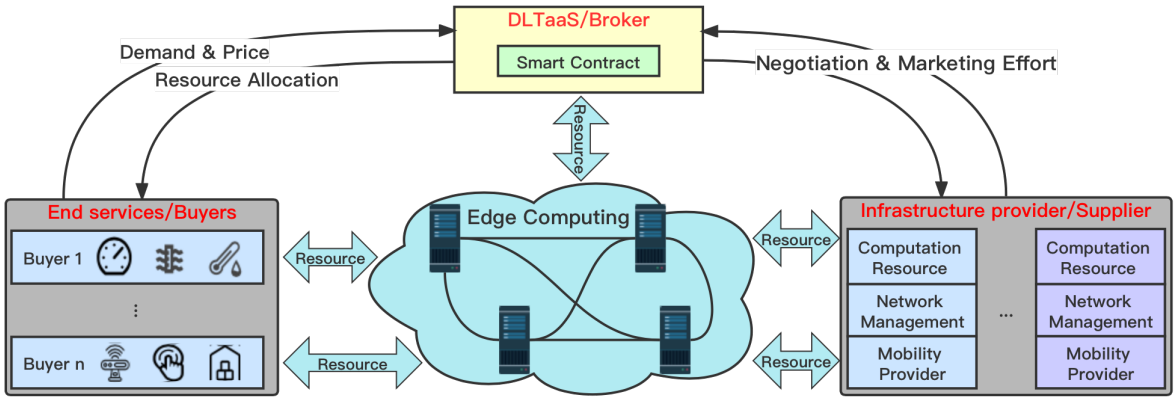


Fig. 1. Architecture of the DLT assisted IIoT marketplace

conclusions in Section VIII.

II. Related work

This section presents related works that consider blockchain/DLT assisted IoT marketplaces (since the topic of IIoT marketplaces is limited) and their pricing strategies as a whole. Reinforcement learning has become popular in solving resource allocation/pricing problems in a DLT assisted IoT market. Asheralieva et al. [13] adopted blockchain as an IoT data management solution to overcome limited scalability, single point failure, and lack of transparency. This paper used hierarchical deep learning to perform distributed dynamic resource management and a pricing strategy between the mobile edge computing servers and the blockchain peers. The work in [14] resolved a similar problem to [13] by using an asynchronous advantage actor-critic (A3C) deep reinforcement learning algorithm, which resulted in extra communication cost between the actor and the critic. Yao et al. [15] studied the resource trading problem between the cloud provider and the miners in a blockchain-based industrial Internet of Things. That paper utilized a multi-agent reinforcement learning algorithm to achieve the Nash Equilibrium.

It is typical to price the resources and services in IoT following game theoretic approaches [16]. In [17], [18], a two-stage Stackelberg game was proposed to determine the optimal price between the consumers and owners in the DLT assisted IoT market. Hu. et al. [19] present a Blockchain-based reward mechanism for mobile crowdsensing (MCS). It suggested a three-stage Stackelberg game to facilitate the reward scheme among the monthly-pay, instant-pay participant, and the task initiator. Due to the selfishness of the players in the market, they may try to hide information to obtain a higher payoff while interacting with others, this is termed a *moral hazard*. The works in [20], [21] used contract-theoretic pricing strategies to tackle the moral hazard between the IoT users and the blockchain.

To summarize, [13]–[15] studied the possibility of utilizing reinforcement learning to solve the resource allocation and pricing problem. but these solutions had several limitations. First, deep reinforcement learning requires a longer training time compared to Q-learning, which renders it inappropriate for dynamic IoT marketplaces as are studied here. This is because retraining will happen frequently. Second, multi-agent reinforcement learning and actor-critic approaches require additional communication costs and impose unnecessary delays (or may result in training with outdated information). Third, although DLT is computation-intensive and consumes resources in the IoT market, it also provides valuable security services to the IoT market, which others can purchase. For approaches based on game theory [17]–[21], it is challenging for them to capture the market dynamics when the number of participants increases. They need to recalculate whenever there is a new demand from each participant. In addition, the participant cannot have complete information about the market.

Differing from existing approaches, in this paper:

1) we propose an online approach based on Q-learning with independent learners, which reduces the training time compared to centralized learning and deep reinforcement learning approaches;
2) we follow an independent learner approach to avoid communication between the buyers. This helps to preserve the private information and reduce the communication cost. Further, the proposed algorithm captures the market dynamics and provides a fast response to them.
3) we employ the Hashgraph based consensus mechanism and consider the computation cost of the transaction verification, which enables micropayment in IIoT scenarios.
4) the DLT becomes a resource that the end services want to purchase to assist trading security.

The works that study architecture and resource optimizations of blockchain/DLT assisted IoT [9], [22]–[25] are not directly applicable to solve the problems considered in this paper.

III. Architecture of the marketplace

Here, we present the proposed architecture of the DLT assisted IIoT market (Fig. 1) and define the roles and interactions of the entities, i.e., buyers, supplier, and the broker. We focus on the costs required to promote sales, namely the marketing effort of the supplier and the broker. We, then, introduce the

broker who operates a Hashgraph-based consensus mechanism to coordinate the supplier and the buyer while providing the necessary security requirements of the marketplace. The broker should be a third party (besides the resource provider and end services provider) to guarantee the fairness of the market. This third party can be governmental agencies or profitable institutes. The revenue will be shared between the supplier and the broker while operating the marketplace, since they both contribute valuable resources and services.

A. Players in the marketplace

The supplier has a number of resource combinations to sell which can be: computation resources, communication resources, and other management resources. To enable fast resource allocation in IIoT applications, we assume the supplier hosts edge computing resources to support the buyers' service demand, as shown in Fig. 1. According to the buyers' demand, the corresponding resource will be allocated.

The broker (i.e., DLTaaS provider) is placed between the supplier and the buyers in the supply chain. It is not only mediating interactions between the two, but also it provides a fair and secure service to both the supplier and the buyers through a hashgraph-based consensus mechanism. In order to operate the consensus mechanism, the broker requires computation resources. In this paper, we assume the edge computing can provide the required resources to the blockchain peers operated by the broker. In the real scenario, the computing resources can be edge, fog, and cloud computing. The supplier and the broker form a contract of the *revenue share ratio* according to the sales of the resource, security service, and their marketing efforts. As the DLT mediates between the supplier and the buyers, it obtains information from both these parties.

The buyers make decisions on the amount and combination of infrastructure resource to buy and the unit price according to resource's reputation. Resource reputation is positively related to the demand quantity, as is formally defined later in Section IV-C. As expected the buyers are more satisfied when they can estimate the reputation accurately. It is possible to obtain the resource reputation through professional product review or historical sales record [26]. However, the information is difficult to rely on, due to such factors as: lack of information transparency and biased or even falsified information. Moreover it is not only hard to acquire the resource reputation, but also the buyers need to share the limited resource and service with other buyers, which adds a further level of difficulty in the buyers' decision-making process.

We assume that the buyers and the supplier have legitimate authenticated identities before entering the marketplace, for example through a policy-based framework as proposed by the SerIoT architecture [7], [27].

B. Consensus mechanism of Hashgraph

In the proposed work, the consensus mechanism is used to not only verify the identities of the stakeholders, but also the transactions they generate during trading.

Hashgraph adopts the gossip about gossip protocol [8] to form a DAG. After the DAG is formed, every *full node* runs traditional Byzantine Fault Tolerance locally, which is termed *virtual voting*. A full node has the complete history of the hashgraph and, thus, has the full voting information. Therefore, it can perform virtual voting without further communication. Thus, virtual voting only costs a little local computation and communication resources compared to the PBFT consensus mechanism [28].

Fig. 3 (i) shows four full nodes, namely Alice, Bob, Carlo, and David. Bob creates an event that has the structure shown in Fig. 2. Then, Bob signs and sends it randomly to a node, David in this case. David inspects the event, and creates a new event with the transactions Bob does not know. David then signs the event with its signature, and sends the new event randomly, in this case, to Bob. This *gossip about gossip protocol* goes on indefinitely creating the DAG. Following this protocol, soon all the full nodes will have a copy of the Hashgraph, which allows them to run the consensus mechanism locally.

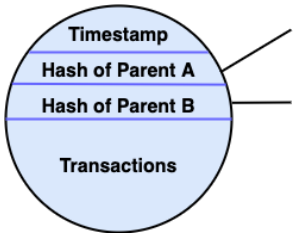


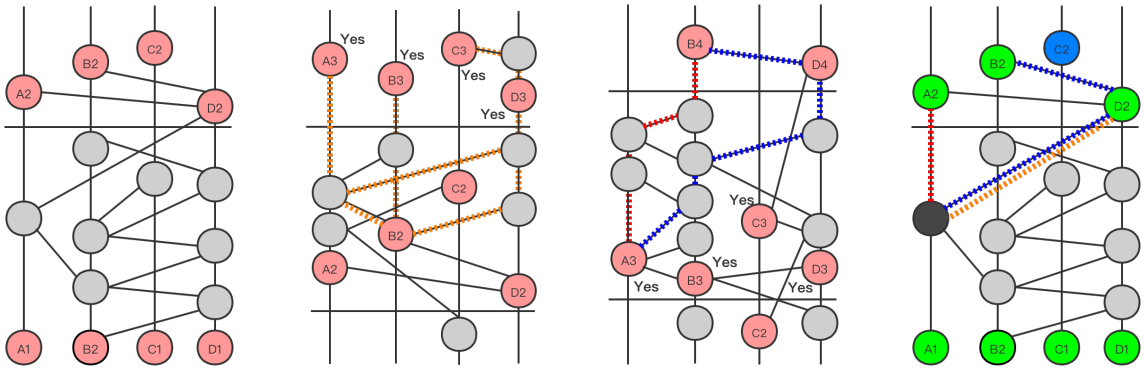
Fig. 2. Event structure of Hashgraph: including timestamp, hash of two parents, and the transactions.

To verify the transactions, the full nodes conduct *virtual voting*. The first event in each round is called a witness. If the witness in round $x+1$ sees the witness in round x, then it votes "yes". For example, in Fig. 3 (ii), $B2$ is seen by witnesses in round 3, i.e., $A3$, $B3$, $C3$, and $D3$ vote "yes" through the orange dashed line. The vote goes on until all the witnesses in round $x+1$ finish voting. Since the full nodes have a copy of the whole Hashgraph, the voting procedure is actually performed locally without any communication cost. The votes are *counted* by witnesses in round $x+2$. Witnesses count the vote, only if it *strong sees*, i.e., there are different paths across a supermajority of population (more than two thirds of the population). For example, in Fig. 3 (iii), there are only witnesses $B4$ and $D4$ till now. $B4$ counts the vote "yes" of $A3$ through the red and blue dashed line, which goes across Alice, Bob, and David (i.e., satisfies supermajority). When the witness from $x+2$ round collects "yes" from a supermajority, then we say the witness from x round is *famous*, which means $B2$ is famous. Note that there is a *coin round* every ten rounds of voting to make sure the election would finally end. At the end, the events are ordered according to the votes, as shown in Fig. 3 (iv).

C. Security benefits of the architecture

While the motivation for an online marketplace is highlighted in the Introduction (with more depth given in such texts as [4]) the security of such a marketplace is essential. Additionally, a central part of the marketplace is to provide *fairness* which is a central tenet that is widely agreed [4].

This article has been accepted for publication in IEEE Transactions on Network and Service Management. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TNSM.2022.3191979



(i) Hashgraph (ii) Famous witeness and vote (iii) Count vote (iv) Order

Fig. 3. Consensus mechanism of Hashgraph [8] representing nodes Alice, Bob, Carlo and David and their associated events (A2, A3, B2 etc.).

Thus, it is vital that entities involved in the marketplace can inspect the information, in particular reputation information, to ensure that it is accurate and transparent. Thus, a DLT is the natural choice as it provides immutability of the data such that entities can inspect it to ensure that fairness has been applied in the past, and thus have confidence that it will be applied in the future. This gives incentives for the broker to act fairly as failure to do so will be seen in the ledger. The general security of DLT has been widely covered in the literature and will not be further covered here, see the many sources on this topic in a review paper such as [29], [30]. The choice of Hashgraph [31] has some security advantages over some other DLT solutions, for example its use of asynchronous Byzantine fault tolerance means that no single entity (or small group of entities) can act maliciously, additionally it provides Byzantine fault tolerance in its *strongest* sense. For example, blockchain does not provide Byzantine fault tolerance as it instead provides a probabilistic approach to consensus, whereas Hashgraph provides strong guarantees of consensus knowledge in all nodes. However, while Hashgraph provides some strong motivations for security, in this paper the strongest motivation for Hashgraph is that it uses a Byzantine consensus mechanism that is highly efficient compared to *proof-of-work* consensus which is so inefficient it is damaging to the planet's ecosystem [32].

IV. System Model

We model the marketplace as a system of: *suppliers* (infrastructure providers), *brokers* (DLTaaS) and *buyers* (end service providers). The supplier offers to the buyers, via the broker, a selection of resources. The supplier also agrees with the broker on a *revenue share ratio* ϕ. The supplier makes effort e^s to promote the sales of resources with a cost function denoted as $C^s(e^s)$. The broker also promotes the sales of resources on behalf of the suppliers and aims to provide fairness across different suppliers and buyers. This is achieved through execution of a *consensus* service. We denote the broker's effort as e^b with a cost function $C^b(e^b)$. There is a group of buyers $i \in \mathcal{M} = \{1, ..., M\}$ in the IoT marketplace. Before joining the market, the buyers observe the market and gather information aiming to estimate the resource reputation θ. Then, buyer i proposes a unit price p^i according to the market demand d^i, which is related to the resource reputation. Furthermore, each of the players has a utility function that reflects their satisfaction. Next, we model the utilities and the *social welfare* (collective utility value) in an IoT marketplace. The most important notations are summarized in Table I.

A. The supplier's utility

Before trading starts, the supplier needs to agree a fair revenue sharing ratio ϕ with the broker, which incentives the latter to make an effort to promote the sale of resources. We denote the income of the supplier as $\mathcal{I}^s$. The net profit of the supplier is therefore the income minus the promotion cost of the supplier and the revenue share paid to the broker, formulated as:

$$U^s = (1 - \phi)\mathcal{I}^s - C^s(e^s), \tag{1}$$

where the marketing effort is an integer variable related to the particular resource combination type that the supplier is providing to the buyer. Note that, the utility function represents the satisfaction level of the supplier and needs to be positive.

B. The broker's utility

In Hashgraph, the gossip about gossip protocol is deployed for transaction dissemination. We assume that each full node randomly selects another full node to transmit a new transaction to. During the transmission, the new transaction is forwarded from the source to the destination without other information exchange. For this new transaction to be known by all the N full nodes, it should be transmitted at least $N - 1$ times. We define μ as the probability of a full node being famous (who has the voting rights) and voting for this transaction. The full node creates, signs, and sends the event. Then, the next full node who receives it will inspect and create a new event (along with the transactions it has), sign, and send to another node. We denote the CPU per event inspection,

creation, signing, and send as δ, κ, ξ, and υ, respectively. We define the cost of one event creation as

$$e^b = \lceil \delta + \kappa + \xi + \upsilon \rceil \tag{2}$$

Due to the fact the edge nodes running DLT have limited computational resource to provide in the IoT marketplace (i.e., there may be other tasks running on the servers), we assume there are only limited security and fairness levels the broker can make. Thus, we round up the marketing effort of the broker. In Hashgraph, voting, counting, and ordering run locally in the famous full nodes. We denote the cost function of the broker in two parts: the new transaction dissemination cost, and the voting and ordering cost. Thus, we have the CPU requirement for one transaction as

$$C^b(e^b) = (N-1)e^b + \mu N {e^b}^2, \tag{3}$$

where $(N-1)e^b$ is the transaction dissemination cost among $N-1$ full nodes and $\mu N {e^b}^2$ represents the higher voting and ordering cost. In this paper, we are aware that the voting and ordering cost is related to the dissemination cost due to the tasks running on the same machine. Thus, we assume that the voting and ordering cost is a quadratic function of the dissemination cost, which is similar to the way cost is defined in [19]. We should note that other formulations are possible but they do not change the process of the derivations below.

The broker provides additional value to the infrastructure resource, i.e., providing a secure marketplace, and in return the broker takes a revenue share from the supplier. We define the utility of broker including the income of selling the resource $\mathcal{I}^s$ and the effort cost as

$$U^b = \phi \mathcal{I}^s - C^b(e^b). \tag{4}$$

C. The buyers' utility

Buyer i's demand is promoted by the efforts of the supplier and the broker. At the same time, the demand is also related to buyer i's estimated reputation of the resource, θ^i. Thus, we can define the demand of the buyer i as

$$d^i = \theta^i + \alpha^i e^s + \beta^i e^b \tag{5}$$

where α^i and β^i are positive perception parameters of the marketing efforts e^s and e^b, respectively. Note that, the supplier and broker can only supply the IoT marketplace with limited resource D. We denote the unit price of the resource combination as p^i. The buyer i's utility function is given by

$$U^i = \eta \theta^i - p^i d^i \tag{6}$$

where the income of buyer relates to the resource reputation. $\eta > 0$ is the preference of the resource reputation, i.e., the bigger is the preference factor, the more the buyer prefers this resource. The buyer aims to purchase more high reputation resource than low reputation resource, i.e., good quality of the resource leads to satisfaction. Since there is only limited resource provided by the supplier and the broker, the buyers' resource demands satisfies

$$\sum_{i=1}^{M} d^i \leq D \tag{7}$$

D. Social welfare in IIoT marketplace

In the IoT marketplace, the *joint satisfaction* of the entities is an important factor, namely the sum of the utilities of the supplier, broker, and the buyers.

$$U^{sw} = U^s + U^b + \sum_{i=1}^{M} U^i \tag{8}$$

In this work, all the entities are rational, which means the utility functions need to satisfy: $U^s \geq 0$, $U^b \geq 0$, and $U^i \geq 0, i \in \mathcal{M}$.

According to the IoT marketplace, we discover that the accuracy of the estimated resource reputation affects the market demand of the buyer. Furthermore, it affects the marketing efforts of the supplier and the broker. In this paper, we first propose an offline approach in Section V. The proposed analytical approach assumes the resource reputation is constant over time which allows the problem to be cast as a linear optimization problem. Then, we take one step further for the time varying scenario and model the problem as a POMDP. We solve it using Q-learning, which allows us to remove the assumptions made in the analytical approach in Section VI-B as well as to provide an online algorithm that allows us to deal with the more realistic case where the resource reputation varies with time.

TABLE I. Descriptions and notation

Description	Notation
A group of buyers	$i \in \mathcal{M} = \{1, ..., M\}$
Resource demand	d
Marketing effort of supplier	e^s
Number of full nodes in DLT	N
Marketing effort of broker	e^b
Probability of being famous	μ
Revenue sharing ratio	ϕ
Reputation of the resource	θ^i
Unit price of the resource	p^i
Weighting of broker's effort	α^i
Weighting of supplier's effort	β^i
Weighting of buyer's income	η
State space	$\mathcal{S} = \{s_0, s_1, ...s_I\}$
Observable state space	$\mathcal{O} = \{o_0, ..., o_I\}$
Action space	$\mathcal{A} = \{a_0, ..., a_J\}$
Probability of observing o_i at state s_i	$P(o_i \vert s_i)$
Transition probability of state s_i	$P(s_{i+1} \vert s_i)$
Immediate reward of state i action j	r_{ij}
Discount factor	γ
Learning rate	α
Greedy factor	ϵ

V. Analytical problem formulation

As analyzed in the last section, we can intuitively formulate the studied problem into an optimization problem. Note that, in the objective function there are a mixture of integer parameters (*e.g.,* marketing efforts) and continuous variables thus the optimization problem is a mixed-integer linear program (MILP). To simplify the problem and clearly show how the efforts would affect the solution of the problem, we first assume the resource reputation is constant and then relax the

integer constraint (marketing efforts). In this section, we use the subscripts is or ia to represent the information symmetric and asymmetric cases, respectively. We discuss the information symmetric scenario first in Section V-A, where the broker and the supplier know each other's marketing effort, which is an ideal scenario. In addition, we explore an information asymmetric scenario, where the supplier and the broker hide each other's marketing efforts in Section V-B.

A. Information symmetry

We first assume that the supplier and the broker know each other's marketing effort which, in other words, corresponds to an *information symmetric* scenario. In this scenario, we say the problem is ideal and will reach the optimal utility and marketing effort without extra cost compared to the moral hazard scenario we consider later. According to the proposed supply chain, we apply backward induction to solve the optimisation problem. First, we assume the unit price p^i of buyer i is given to solve the supplier optimization problem. Then, we consider the optimization problem of the buyers. The objective of this problem is to maximize the utility of the supplier. We cast the optimization problem as follows

$$\max_{\phi_{is}, e^s_{is}, e^b_{is}} \quad (1-\phi_{is}) \sum_{i=0}^{M} p^i_{is} d^i_{is} - {e^s_{is}}^2 \tag{9a}$$

$$\text{s.t.}$$

$$\phi_{is} \sum_{i=0}^{M} p^i_{is} d^i_{is} - [(N-1)e^b_{is} + \mu N {e^b_{is}}^2] \geq 0 \quad \text{(IR)} \tag{9b}$$

where the income of the supplier is the product of the unit price of resource p_{is} and the resource demand d^i_{is}. The income of the broker considers the revenue sharing ratio with the supplier. The constraint in (9b) is the *Individual Rationality* (IR) constraint of the broker. We can obtain the revenue share ratio ϕ when IR is equal to zero

$$\phi_{is} = \sum_{i=0}^{M} \frac{1}{p^i_{is} d^i_{is}} [(N-1)e^b_{is} + \mu N {e^b_{is}}^2] \tag{10}$$

Then, we first substitute (10) into (9a), and obtain

$$\max_{e^s_{is}, e^b_{is}} U^s_{is} = \sum_{i=0}^{M} p^i_{is} d^i_{is} - [(N-1)e^b_{is} + \mu N {e^b_{is}}^2] - {e^s_{is}}^2 \tag{11}$$

We then substitute the market demand in (11) considering the market demand as defined in (5). In order to calculate the optimal efforts, we set both the partial derivatives of the utility function with respect to e^s_{is} and e^b_{is} in (11) to zero as follows:

$$\begin{aligned} \frac{\partial U^s_{is}}{\partial e^s_{is}} &= \sum_{i=0}^{M} p^i_{is} \alpha^i - 2e^s_{is} = 0 \\ \frac{\partial U^s_{is}}{\partial e^b_{is}} &= \sum_{i=0}^{M} p^i_{is} \beta^i - [(N-1) + 2\mu N e^b_{is}] = 0 \end{aligned} \tag{12}$$

Then, we obtain the optimal efforts of the supplier and broker, respectively, as

$$\begin{aligned} {e^s_{is}}^* &= \frac{1}{2} \sum_{i=0}^{M} p^i_{is} \alpha^i \\ {e^b_{is}}^* &= \frac{1}{2\mu N} [\sum_{i=0}^{M} p^i_{is} \beta^i - (N-1)] \end{aligned} \tag{13}$$

In the information symmetric scenario, the supplier chooses the effort to maximize its utility and the total utility of the supply chain. Here, the contract, namely the revenue sharing ratio can be directly constructed according to the effort of the broker as is in (10), in which case the revenue sharing ratio can compensate the cost of the broker. We now solve the buyers' optimization problem. According to (13), the marketing efforts are related with the unit price of all the buyers. Hence, we analyze the payoff of the buyers as a whole.

$$\max_{p^i_{is}} \quad \sum_{i=1}^{M} U^i_{is} = \sum_{i=1}^{M} [\eta \theta^i - p^i_{is} d^i_{is}] \tag{14a}$$

$$\text{s.t.}$$

$$\sum_{i=1}^{M} d^i \leq D \tag{14b}$$

We substitute the optimal marketing efforts ${e^s_{is}}^*$ and ${e^b_{is}}^*$ from (13) into (14a) and obtain the derivative of unit price p^i.

$${p^i_{is}}^* = \frac{\beta_i (N-1) - 2\theta^i \mu N}{2\mu N {\alpha^i}^2 + 2{\beta^i}^2} \tag{15}$$

B. Information asymmetry

The information symmetric case assumes sharing of marketing effort, which is unlikely in practice. The result of hiding the real marketing efforts causes the information between the supplier and the broker to be asymmetric, e.g., corresponds to a moral hazard scenario. The main issue with the moral hazard is that both the supplier and the broker cannot exert optimal marketing efforts according to each other's behavior. Following the solution structure proposed in Section V-A, we cast the optimization problem of the supplier as

$$\max_{\phi_{ia}, e^s_{ia}} \quad U^s_{ia} = (1-\phi_{ia}) \sum_{i=0}^{M} p^i_{ia} d^i_{ia} - {e^s_{ia}}^2 \tag{16a}$$

$$\text{s.t.}$$

$$\phi_{ia} \sum_{i=0}^{M} p^i_{ia} d^i_{ia} - [(N-1)e^b_{ia} + \mu N {e^b_{ia}}^2] \geq 0 \tag{16b}$$

where the optimization variables for the supplier are e^s_{ia} and ϕ_{ia}. The utility of the broker is positive given the revenue sharing ratio ϕ from the supplier. The maximization problem solved by the broker is

$$\max_{e^b_{ia}} U^b_{ia} = \phi_{ia} \sum_{i=0}^{M} p^i_{ia} d^i_{ia} - [(N-1)e^b_{ia} + \mu N {e^b_{ia}}^2] \tag{17}$$

where $e^b_{ia} > 0$ and $U^b_{ia} > 0$. In (16a) and (17), the supplier and the broker aim to maximize their utilities individually

due to the moral hazard. We consider this as a two stage problem: in the first stage the supplier designs a contract (i.e., revenue sharing ratio) with the broker including its own effort to maximize the utility in (16a); in the second stage the broker takes the contract from the supplier and exerts its own effort to maximize the utility in (17). To solve this problem, the supplier and the broker's marketing efforts are obtained by the partial derivatives of (16a) and (17), respectively as shown below

$$\frac{\partial U^s_{ia}}{\partial e^s_{ia}} = (1-\phi_{ia})\sum_{i=1}^{M} p^i_{ia}\alpha_i - 2e^s_{ia} = 0 \tag{18}$$

$$\frac{\partial U^b_{ia}}{\partial e^b_{ia}} = \phi_{ia}\sum_{i=1}^{M} p^i_{ia}\beta_i - [(N-1) + 2\mu N e^b_{ia}] = 0 \tag{19}$$

According to (18) and (19), ϕ_{ia} can be represented by

$$\phi_{ia} = \sum_{i=1}^{M} \frac{\alpha_i[(N-1) + 2\mu N e^b_{ia}]}{\alpha_i[(N-1) + 2\mu N e^b_{ia}] + 2\beta_i e^s_{ia}} \tag{20}$$

Recalling that $(N-1) + 2\mu N e^b_{ia} = C^{b\prime}(e^b_{ia})$ and $2e^s_{ia} = C^{s\prime}(e^s_{ia})$ then, (20) relates the revenue share ratio with the marking efforts of the supplier and the broker. (20) requires the supplier and the broker to choose their market efforts so that the ratio of the broker's marginal cost to the total of marginal cost is equal to the revenue sharing ratio. Then, the supplier chooses the optimal revenue sharing ratio by conducting the partial derivative of ϕ_{ia} in equation (16a) as shown below

$$\frac{\partial U^s_{ia}}{\partial \phi_{ia}} = -\sum_{i=1}^{M} p^i_{ia} d^i_{ia} + (1-\phi_{ia})\sum_{i=1}^{M} p^i_{ia} {d^i_{ia}}' = 0$$

$$\phi_{ia} = 1 - \sum_{i=1}^{M} \frac{\theta^i + \alpha^i e^s_{ia} + \beta^i e^b_{ia}}{\frac{\partial e^b_{ia}}{\partial \phi_{ia}}} \tag{21}$$

Due to broker's marketing effort being related with the revenue sharing ratio, we have e^b_{ia} to be a function of ϕ_{ia}. We can solve $\frac{\partial e^b_{ia}}{\partial \phi_{ia}} = \frac{\sum_{i=1}^{M} p^i_{ia}\beta^i}{2\mu N}$ according to (19) and obtain the revenue sharing ratio

$$\phi_{ia} = \sum_{i=1}^{M} \frac{1}{2A}[A - p^i_{ia}\theta^i - p^i_{ia}\alpha^i e^s_{ia} + B] \tag{22}$$

where $A = \frac{\sum_{i=1}^{M} {\beta^i}^2 {p^i_{ia}}^2}{2\mu N}$, $B = \frac{\sum_{i=1}^{M} p^i_{ia}\beta^i}{2\mu N}(N-1)$. From (20), we know that in the IA scenario, the supplier can only design the contract according to its own effort. Finally, we substitute the optimal effort of broker from (19) and the revenue sharing ratio (20) into the utility function of the supplier in (16a) to obtain the optimal value of the supplier's effort

$${e^s_{ia}}^* = \frac{1}{2}(1-\phi_{ia})\sum_{i=1}^{M} p^i_{ia}\alpha^i$$

$${e^b_{ia}}^* = \frac{1}{2\mu N}[\phi_{ia}\sum_{i=1}^{M} p^i_{ia}\beta^i - (N-1)] \tag{23}$$

The cost functions of the supplier and the broker are increasing and convex functions, as defined earlier, so we find that the first derivatives of the cost function are $C^{s\prime}, C^{b\prime} > 0$ and the second derivatives are $C^{s\prime\prime}, C^{b\prime\prime} \geq 0$. We now solve the buyer i's optimization problem following the method in V-A

$$\max_{p^i_{ia}} U^i_{ia} = \eta\theta^i - p^i_{ia} d^i \tag{24}$$

We substitute the optimal marketing efforts ${e^s_{ia}}^*$ and ${e^b_{ia}}^*$ from (23) into (24) and obtain the derivative of unit price p^i_{ia}.

$${p^i_{ia}}^* = \frac{\beta_i(N-1)}{2\mu N {\alpha^i}^2(1-\phi_{ia}) + 2\phi{\beta^i}^2} \tag{25}$$

Lemma 1. *When $0 < \phi < 1$, we have ${e^s_{ia}}^* < {e^s_{is}}^*$ and ${e^b_{ia}}^* < {e^b_{is}}^*$, if and only if $C^{s\prime}, C^{b\prime} > 0$ and $C^{s\prime\prime}, C^{b\prime\prime} \geq 0$.*

Proof. This lemma is a direct application of (13) and (23). □

As this is the information asymmetric scenario, first, the efforts are smaller than that in the information symmetry scenario; second, the revenue sharing ratio is related to both the efforts of the supplier and the broker. The supplier designs a contract to encourage the broker to make more effort. Hence, we can interpret the difference between supplier's efforts in IS and IA as the *Information acquisition Cost* (IC). We define the information acquisition cost as a function $IC(\cdot)$. We can obtain the IC of the supplier $IC({e^s_{is}}^* - {e^s_{ia}}^*)$, which is related to the the effort perception parameters α and β from buyer M. Similarly, the IC of buyer i is $IC(|{p^i_{is}}^* - {p^s_{ia}}^*|)$, which is related to the performance parameters of the DLT and the revenue share ratio from the supplier. However, the revenue share ratio is often a business secret, which leads to buyers' $IC(\infty)$.

In the analytical solutions, we first assume the resource reputation θ is known, however, *the resource reputation is never transparent to the buyer*, which leads to the resource demand fluctuating with the resource reputation. Second, *the marketplace is dynamic*, namely in each trading period the buyers' demand and the marketing efforts influence the marketplace, which results to uncertainty in the next trading period. Third, *the buyers are heterogeneous* with varying demands, perceptions, and bidding strategies (e.g., risk averse and risk seeking). Additionally, for each trading period, we need to re-calculate a new strategy, which makes this offline approach suffer from high latency. This necessitates the design of an online approach with independent learning agents which we present in the following section.

VI. POMDP WITH INDEPENDENT LEARNERS

The learning approach aims to capture the supply and demand relationship dynamics of the proposed DLT-assisted IoT resource market, which allows us to find the equilibrium of the unit resource price and the resource quantity. Hence, we deploy independent Q-learning agents that run a reinforcement learning algorithm to decide the optimal actions considering the market dynamics. The independent learning agents, that represent the buyers, are empowered by the edge computing resource. The agents can observe partial information of the IoT marketplace. The buyers analyze the resource reputation of the supplier and the service reputation of the broker and then independently update their belief of the value of the resource. According to this belief, the buyers would decide the optimal

This article has been accepted for publication in IEEE Transactions on Network and Service Management. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TNSM.2022.3191979

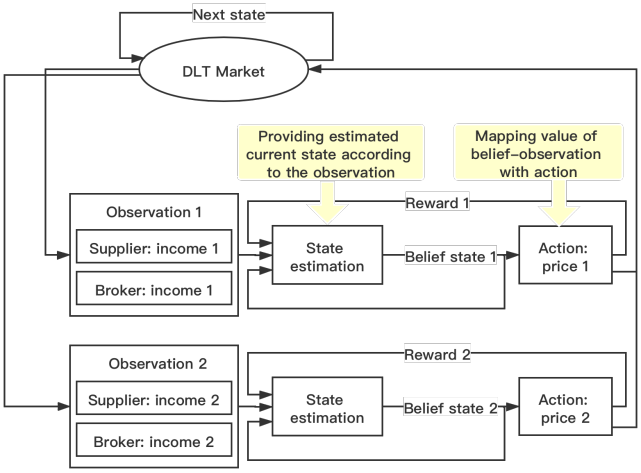


Fig. 4. Decision process following POMDP framework.

unit price. To employ reinforcement learning, we first model the above procedure as a Markov Decision Process (MDP).

In Fig. 4, we illustrate the flow diagram of the decision-making process of two independent agents that interact with the *DLT-assisted market*. We model the marketplace as a POMDP due to the information asymmetries in the market. The POMDP aims at modeling the optimal actions of buyers, and unlike the previous case is appropriate for a non-static scenario. Next, we present the state, action, immediate reward, and transition probability of the considered POMDP.

A. Modeling the POMDP

In the DLT marketplace, the state of the MDP, $S \in \mathcal{S}$, comprises the state of the market S^M and the state of each agent $S^i, i \in \{1, ..., M\}$. Agent i could only observe the state of the market and the state of itself, but not other agent's state, which leads to partial observation of environment. We define the *finite state space* as

$$\mathcal{S} = S^i_{\times 1 \leq i \leq M} \times S^M \tag{26}$$

where the state of buyer i, S^i, is the resource demand d^i and the state of the market, S^M, is the income of the broker and the supplier, i.e., $\mathcal{I}^i_b$ and $\mathcal{I}^i_s$. Each agent obtains the income information from the supplier and the broker's annual report. Hence, the observation state of agent i is $o^i = \{\mathcal{I}^i_b, \mathcal{I}^i_s\} \in \mathcal{O}$, where $\mathcal{O}$ is the set of all observations. The agent aims to anticipate the resource's reputation, the marketing efforts, which is defined as belief state $b(S)$.

At each buyer, there is located an independent agent who learns from the DLT market and makes the price choice for the market demand in order to optimise the payoff of the buyer. According to the current observations and the market demand, agent i takes an action of the unit price from a set of actions $A^i = \{a^i_1, ..., a^i_J\}, a^i_j \in A^i$. Every buyer has independent action set. We then define the action space $\mathcal{A} = \{A^1, ..., A^M\}$, where $A^i \in \mathcal{A}$ is the action set of buyer i.

The transitions from one state to another are determined by the transition probability $p(S_{t+1}|S_t, \mathcal{A}_t)$, which determines the transition to a state at time slot $t+1$ given the state and action at time slot t. Each agent takes an action according to the current state individually. The joint action of the agents influences the next state of the DLT market. This means that the joint action of the unit price affects the marketing efforts of the broker and the supplier, which in turn affects the total resource quantity in the market. In the description of the POMDP below, we drop the time index for the sake of simplicity of representation. In order to calculate the transition probability according to agent i's belief state $b(S^i)$ we follow the procedure below. Let us first define the observation probability as $P(o^i|s^i)$, which is the probability of observing o^i while being in the state s^i. The agent updates the belief state individually as the learning procedure continues as follows

$$b'(S') \propto P(o|S') \sum_{S \in \mathcal{S}} P(S'|A, S) b(S) \tag{27}$$

where the probability to be in new belief state $b'(S')$ is proportional to the product of the probability of observing o while being at state S' and the sum of probabilities of all the states and actions A that the lead to the new state S'. Finally, we can define the transition probability of the POMDP as

$$P(b'|A, b) = \sum_{o \in \mathcal{O}} P(b'|o, A, b) \sum_{S' \in \mathcal{S}} P(o|S') \sum_{S \in \mathcal{S}} P(S'|S, A) b(S) \tag{28}$$

The transition probability matrix (matrix containing the transition probability for all possible state-action combinations) is difficult to calculate due to the size of the action and state space. To address this problem, we adopt Q-learning [33], which does not require calculation of this matrix.

In the POMDP with decentralised agents, agent i defines and evaluates its policy, i.e., the state-action pair, via individual immediate reward r_i. It is obvious that the definition of the immediate reward is a critical component of the POMDP. The immediate reward of an agent aims to promote the buyer's satisfaction. If there is utility gain, then the reward is the same as the utility gain. Otherwise, the immediate reward is -1 as a penalty.

Now, we can define the POMDP as the tuple $(\mathcal{S}, \mathcal{A}, P, r_i, \mathcal{O}, P_{o^i})$, where $\mathcal{S}$ is a finite state space, $\mathcal{A}$ is a finite action space, P is the transition probability from current state to the next state, $\mathcal{O}$ is a finite observation space, and P_{o^i} is the observation probability of current state.

B. Q-learning in POMDP with independent learners

In this work, we solve the POMDP problem following the tabular Q-learning approach. Once the Q-table is computed, the optimal action (the one maximizing the reward when being in a state) is determined by checking the Q-table. For decentralised agents, each agent interacts with the DLT market based on individual observation, and updates its own Q-table according to the reward function. It is essential to calculate the long-term expected reward in Q-learning to ensure convergence. To solve the POMDP based on the beliefs MDP framework, we first define the immediate belief reward of belief i and action j for agent i as $\rho(b^i, a^i_j)$ as follows

$$\rho(b^i, a^i_j) = \sum_{S \in \mathcal{S}} b(S) R(S, a^i_j) \tag{29}$$

where $R(S, a_j^i)$ is the reward at state S taking action a_j^i. Then, according to the Bellman equation [34], we compute the belief value $V_t(b^i)$ at time t for agent i under strategy policy $\pi(a_j^i|b^i)$ as

$$V_t(b^i) = \max_{a_j^i \in \mathcal{A}^i} [\rho(b^i, a_j^i) + \gamma \sum_{o^i \in \mathcal{O}} p(o_i|b_i^i, a_j^i) V_{t-1}(b^{i'})] \quad (30)$$

where $\gamma \in (0, 1)$ is the discount factor and ρ is the immediate reward. The discount rate γ represents the importance of the future reward compared to the immediate reward. When γ is close to 1, the agents become farsighted; While when γ is close to 0, the agents act in a myopic way. From the belief value, we can obtain the policy

$$\pi_{ij} = \underset{a_j^i \in A^i}{\operatorname{argmax}} [\rho(b^i, a_j^i) + \gamma \sum_{o^i \in \mathcal{O}} p(o^i|b^i, a_j^i) V_{t-1}(b^{i'})] \quad (31)$$

Each agent aims to maximize the accumulated long-term reward. Instead of belief state value $V_t(b^i)$, we use the Q-value function in the following Q-learning, which provides richer information since the Q-value is a tuple of belief state and action, compared to the belief state value which contains only belief information. Thus, we can obtain the Q-value as following

$$Q(b^i, a_j^i) = E[\sum_{\tau=0}^{\infty} \gamma^\tau \rho(b^i, a_j^i)] \quad (32)$$

Furthermore, we can obtain the optimal policy π^* as follows

$$\pi^* = \underset{a_j^i \in A^i}{\operatorname{argmax}} E[\sum_{\tau=0}^{\infty} \gamma^\tau \sum_{S \in \mathcal{S}} b(S) R\left(S, \pi(b^i)\right)] \quad (33)$$

where $\pi(b^i)$ corresponds to the action taken when in belief state b^i. We then update the Q-value as shown below

$$Q(b^i, a_j^i) \leftarrow (1-\alpha) Q(b^i, a_j^i) + \alpha [\rho(b^i, a_j^i) + \max_{a^{i'} \in A^i} Q(b^{i'}, a^{i'})] \quad (34)$$

where $\alpha \geq 0$ is the *learning rate*. The learning rate indicates the rate that new knowledge is acquired by visiting a new state. As previously discussed, the key to solving this problem is to determine a finite belief space that is not too big to be calculated and not too small to be accurate. To compensate for the infinite belief state, we use a fixed history window to approximate beliefs with a finite-history of observations as in [35]. Note that the choice of a fixed history window affects the size of the Q-table (complexity) and the quality of the solution. However, its thorough investigation is out of the scope of this paper.

C. Algorithm and its complexity

In the proposed work, we have defined the action space as the unit price of the buyers. The size of the action space of agent i is $|A^i|$. To simplify the problem and make it feasible to be solved following tabular Q-learning, let us first consider the utility of the buyer i, recall function where we dropped the time-dependence index (6)

$$U^i = \eta \theta^i - p d^i > 0 \quad (35)$$

Then, we obtain the upper bound of resource demand as

$$p^i < \frac{\eta \theta^i}{d^i} \quad (36)$$

which means that buyer i's action, namely price p^i, is bounded by the reputation and the demand. For the observed state, we note that there exists the minimum and the maximum perceptible production reputation θ_{min} and θ_{max}, respectively, which leads to the price's upper bound as $p^i < \frac{\eta \theta_{max}}{d_{min}}$. The minimum demand d_{min} is reached if the perceptible production reputation stays minimum, and additionally neither the supplier or the broker makes marketing effort. Hence, we obtain the action space of buyer i as $A^i \in [0, \frac{\eta \theta_{max}}{\theta_{min}}]$

For a given resource demand, we can obtain the range of the marketing efforts by keeping the utility functions of the supplier and the broker positive. We substitute the maximum price for the given resource demand d^i, $p^i = \eta \frac{\theta_{max}}{d^i}$ in (1) and (4), then we obtain the ranges of the marketing efforts, $e^s \in [0, x(\phi, \eta, \theta_{max})]$ and $e^b \in [0, y(\phi, N, \mu, \eta, \theta_{max})]$, where $x(\cdot)$ and $y(\cdot)$ are functions of the maximum solutions. Hence, the state space of buyer i is $S^i \in [\theta_{min}, \theta_{max} + \alpha^i e^s_{max} + \beta^i e^b_{max}]$.

Similarly, for the state space of the DLT market, we have the supplier's income and the broker's income as observation states. We can descale the state space according to Section V and constrain the income space as $\mathcal{I}^s = [I^s_{min}, ..., I^s_{max}]$ and $\mathcal{I}^b = [I^b_{min}, ..., I^b_{max}]$. As we mentioned in Section V, the income is shared between the supplier and broker according to revenue sharing ratio ϕ. Furthermore, the income of broker can be represent by $\mathcal{I}^b = \phi \mathcal{I}^s$. As a result, we can reduce the state space by replacing $\mathcal{I}^b$. Thus, the size of states space is $(|\mathcal{I}^s||d^i|)$.

In our framework, the Q-learning algorithm updates a single state-action pair of the Q-table per decision interval according to the received immediate reward. The *computation complexity* is determined by the size of the Q-table, which is $|S| \cdot |A^i| = (|\mathcal{I}^s||d^i|) \cdot (|p^i|)$. Further, the action selection and learning update complexities are equal to $\mathcal{O}(|\mathcal{A}|) = \mathcal{O}(|p^i|)$ [36]. Additionally, as the convergence of the Q-learning is well-understood in [33], we will not repeat the proof of convergence in this paper. We now summarize the Q-learning algorithm with the belief state of the MDP in Algorithm 1. Additionally, we have introduced the information acquisition cost in Section V, where the buyers' IC is the difference between the optimal unit price of IS and IA. We also point out that since the revenue share ratio is a business secret, which leads to $IC(\infty)$ for the buyers. In the online approach, with independent learners who assist the buyers in the decision-making process, the information for decision-making is only related to the income of the supplier and the broker (published in the annual report). The information acquisition cost of the learners is $IC(\mathcal{I}^i_b, \mathcal{I}^i_s)$ when interacting with one supplier and one broker, which is considerably smaller than the analytical solution. When there are multiple suppliers and brokers in the market, the information acquisition cost is bounded by the number of supplier and brokers.

Note the two conditions in Algorithm 1: if "Q did not converge" is evaluated by comparing the performance observed in the last 10, 100, 500 episodes to see if the Q-values are

Algorithm 1: Q-learning with belief state MDP

Data: Initiation: greedy factor ϵ, learning rate α, discount factor γ
Data: Initiation: actions space A, belief space B, and Q-table Q
while Q did not converge **do**
  Select state b arbitrarily;
  while b is not terminal **do**
    Select action a with ϵ-greedy exploitation;
    $r \leftarrow R(b,a)$;
    $b_{next} \leftarrow T(b,a)$;
    Update Q-value:
    $Q(b,a) \leftarrow (1-\alpha)Q(b,a) + \alpha(r + \gamma Q(b,a'))$;
    $b \leftarrow b_{next}$;
  Return Q table;

similar. The termination of the current belief state depends on whether the corresponding action choice leads to a positive reward.

VII. Evaluation

In this section, we evaluate the proposed solution with respect to the achieved social welfare for various marketing efforts and pricing. We have chosen not to carry out a security analysis as this is out of the scope of this work. The security of blockchain solutions has been widely discussed for example by Saad *et al.* [30].

The evaluation starts with the simulation result of the contract design between the supplier and the broker for the case the buyer does not know the resource reputation. Subsequently, we evaluate the contract design with the learning agent by utilizing the Q-learning method. The simulation parameters are as indicated in Table II unless otherwise stated. Based on the complexity in Section VI-C, the computation demand for each learning agent is $(|\mathcal{I}^s||d^i|) \cdot (|p^i|)$ and the information acquisition cost is $IC(\mathcal{I}_b^i, \mathcal{I}_s^i)$. The reward for the learner i is based on the utility function of the buyer i in (6).

TABLE II. Simulation settings of the analytical approach

Description	Setting
Number of full nodes in DLT	$N = 10$
Probability of being famous	$\mu = 0.5$
Reputation of the resource	$\theta = 1$
Weighting of broker's effort	$\alpha' = 2.5$
Weighting of supplier's effort	$\beta' = 15$
Weighting of buyer's income	$\eta = 15$

A. Simulation of IoT marketplace in analytical approach

In this subsection, we evaluate both the information symmetric (IS) scenario, where the supplier designs the contract with the broker by considering its effort directly, and the information asymmetric (IA) scenario, i.e., "moral hazard". By designing a contract, we aim to maximize the joint satisfaction of the supplier, broker, and the buyer. To evaluate the joint satisfaction, we adopt social welfare in (8) as the joint satisfaction of the supply chain.

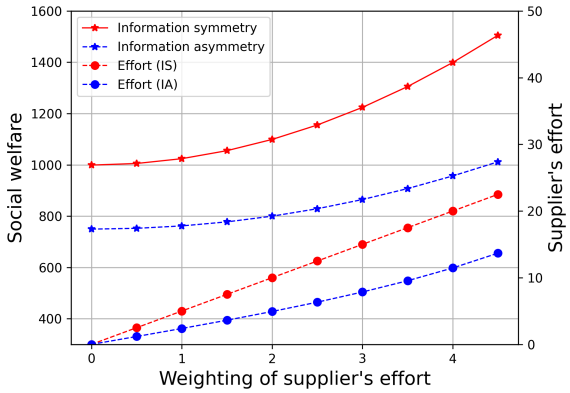


Fig. 5. Weights of supplier's effort α with respect to the social welfare and the supplier's effort e^s

1) Weighting of supplier's effort with respect to the social welfare: First, we investigate the influence of the weights α^i and β^i to the social welfare, which reflect the importance of the marketing efforts. In the simulation, we assume that there is only one buyer in the market. Hence, the superscript of α and β are dropped. As we know, the weights not only relate to the efforts, but also the resource demand d. Fig. 5 illustrates both the effort of supplier and the joint satisfaction/social welfare for IS and IA scenarios. While the weights of the supplier's effort becomes more and more important in the supply chain, i.e., α increases, so does the joint satisfaction. This is due to the fact the supplier's effort significantly influences the market demand, which leads to higher satisfaction of the suppler, i.e., the supplier's utility. This also leads to an increase in the social welfare. The same conclusions can also apply to the weights of the broker's effort β.

2) Marketing efforts with respect to the resource price: We then aim to study the relationship between the efforts in the IS scenario and the IA scenario. By comparing (13) and (23) in Section V, we conclude that the efforts in the IS scenario are only related to the price of the resource and the efforts in the IA scenario are related to both the price and the revenue sharing ratio. Thus, in Fig. 6, we set the resource price to be in the range of 10 to 20 units. Fig. 6 demonstrates that the efforts increase with respect to the price. The reason is that when the price increases, the market demand decreases, which stimulates the supplier and broker to make more marketing effort to promote the sales. Moreover, the efforts of IS surpasses the efforts of IA as is justified through Lemma 1. This is expected as the supplier cannot observe the effort of the broker to design the revenue sharing ratio.

3) Effort of supplier with respect to the social welfare and the revenue sharing ratio: In Fig. 7, we investigate the influence of the satisfaction, i.e., social welfare, and the revenue sharing factor with respect to the supplier's effort. In this simulation, we set the supplier's effort to be in the range of 0 to 40. As we can observe from Fig. 7, the social welfare of IS exceeds that of IA. This is because in IS, the supplier designs the contract (i.e, the revenue sharing ratio) by binding the utility of broker to maximize its utility. The contract of

This article has been accepted for publication in IEEE Transactions on Network and Service Management. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TNSM.2022.3191979

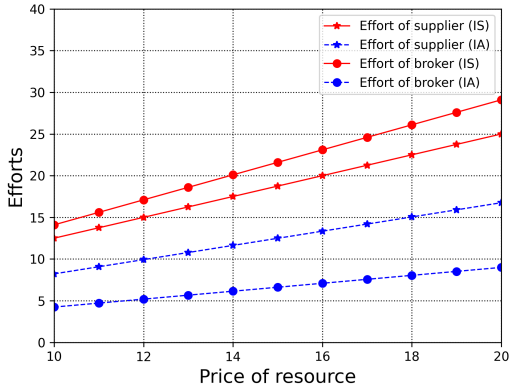


Fig. 6. Supplier's effort e^s and broker's effort e^b with respect to the resource price p

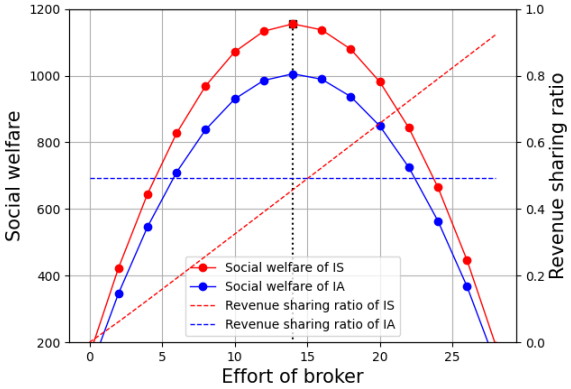


Fig. 8. Broker's effort e^b with respect to the social welfare and the revenue sharing ratio ϕ

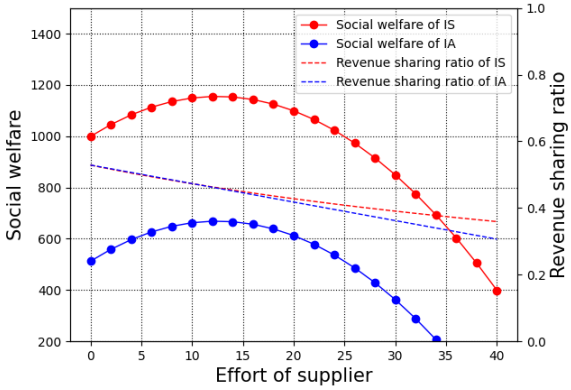


Fig. 7. Supplier's effort e^s with respect to the social welfare and the revenue sharing ratio ϕ

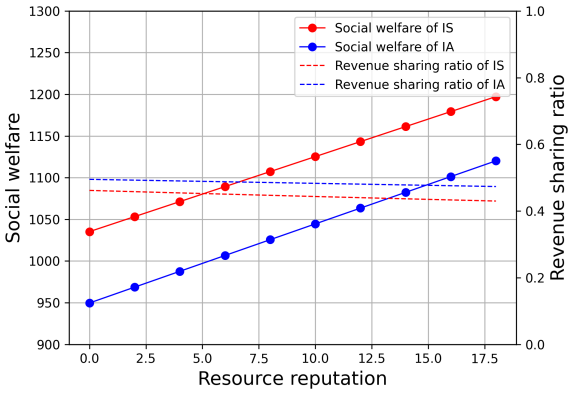


Fig. 9. Resource reputation θ with respect to the social welfare and the revenue sharing ratio ϕ

IS only compensates the cost of the broker. However, in IA the supplier can only design the contract by its own effort and thus motivate the broker to provide more effort. Thus, the incentive model of the IA scenario leads to a satisfaction level lower than the IS scenario. We also explore how the supplier's effort affects the revenue sharing ratio. As is shown in (10) and (22), when the supplier provides increased effort, it shares less revenue with the broker, which leads to a decreasing revenue sharing ratio.

4) Broker's effort with respect to the social welfare and the revenue sharing ratio: Additionally, we investigate the broker's effort with respect to the social welfare and the revenue sharing ratio. In Fig. 8, first we can observe that the joint satisfaction (i.e., social welfare) becomes optimal when the broker changes its effort. The dotted line indicates the relationship with the revenue sharing ratio when the joint satisfaction is at the optimal point. The blacked dotted line has two intersections with the revenues sharing ratio ϕ of the IA first, and then IS, which indicates the ϕ of IA is bigger than IS. This also demonstrates that an additional incentive for the broker is needed.

5) Resource reputation respect to the social welfare and the revenue sharing ratio: In Figs. 6, 7, and 8, we simply fix the reputation of the resource and only focus on the necessity of contract design between the supplier and the broker. Differently, Fig. 9 shows that the higher resource reputation θ leads to higher joint satisfaction. The reason behind this behavior is that the resource demand of the buyer depends on the reputation of the resource, supplier's marketing effort, and the broker's marketing effort. The resource demand is related to the income of the supplier and the broker directly. We should note that in this subsection of the IoT marketplace, where we do not have a learning agent, we boldly assume that the buyer perfectly predicts the resource reputation, which is unlikely to happen in practice. As a consequence, we use the reputation directly for contract design. On the other hand, if we can estimate the resource reputation accurately, the contract design is more efficient in real-life scenarios. To facilitate this point, in the following section, we adopt Q-learning to estimate the resource reputation in the more realistic cases where there are dynamics in the market.

B. Simulation of IoT marketplace with independent learning agents

In this section, we evaluate the proposed Q-learning algorithm in Section VI-B. The evaluation was performed using a complete custom implementation of the Q-learning of Algorithm 1 using Python 3.8. We have reduced the state and action space as shown in the previous section prior to applying the

This article has been accepted for publication in IEEE Transactions on Network and Service Management. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TNSM.2022.3191979

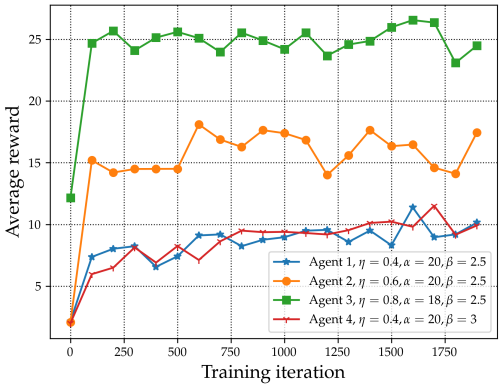


Fig. 10. Performance of q-learning

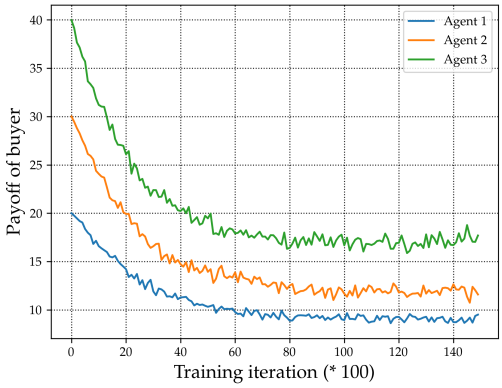


Fig. 11. Policy evaluation with respect to the buyer's payoff

Q-learning algorithm. To optimize the Q-learning algorithm, we dynamically adjust the learning rate α as following

$$\alpha = \frac{k}{k + v(b_i, a_j)}(\log(v+1)+1) \tag{37}$$

where $k > 0$ is offset value and $v(b_i, a_j)$ is the visiting times of the state-action bundle. While the visiting times increase, the learning rate will reduce. We have tune the discount factor and the greedy factor, namely $\gamma = 0.5$ and $\epsilon = 0.95$. We set the history window as 10.

1) Performance of the decentralized approach: We first examine the convergence speed of each independent learning agent with different perceptions of the marketplace, i.e., different weights of the marketing efforts (α and β) and resource value (η). Fig. 10 illustrates that the independent agents require limited training to converge. The resource's importance perception η dominates the average reward of the agents, where the more important agent values the resource, the higher the average reward it would obtain eventually. However, the perception of the marketing efforts has minor influences on the average reward.

2) Optimal policy evaluation: In Fig. 11, we evaluate the best policy that is obtained by the trained model. We deploy three independent agents in the marketplace. This marketplace has a limited market capacity, i.e., the supplier and the broker can only provide a limited resource to the buyers. We first train the model for 30,000 iterations and extract the best action of a certain state, i.e., the best policy. Then, we randomly select a state and execute the best policy 14,000 times. In Fig. 11, we consider the derived Q-table every 100 iterations up to 30,000. The figure illustrates that while the iterations increase, the buyer's payoff reaches an equilibrium, which means that the proposed algorithm can eventually converge.

We observe that the buyer's payoff decreases with the iterations. This shows the buyers' act more and more conservatively, but the action stabilizes in resource bidding due to the limited market capability, which allows the buyers to achieve higher payoff in the market.

3) Scalability of the decentralized approach: Fig. 12 demonstrates the scalability of the decentralized approach with a different number of independent agents. We examine the achieved social welfare. We first observe that in all cases

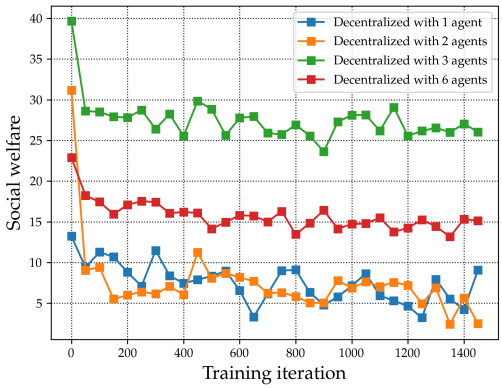


Fig. 12. Scalability of the decentralized approach with multiple learning agents

the social welfare converges, which shows good scalability of the algorithm. Then, we notice that with one or two learning agents in the marketplace, the social welfare is low. However, with more agents in the market, higher social welfare is achieved. This happens as more buyers lead to more competitiveness of the marketplace, which is essential for a healthy market. Finally, for a marketplace with fixed capacity, there exists an optimal number of buyers. In this setting, three buyers reach higher social welfare compared to six buyers due to limited resource supply.

4) Centralized and Decentralized approaches: Last but not least, we compare the performance of the centralized and decentralized approach in Fig. 13. The centralized approach considers one agent that represents all the entities in the marketplace. The centralized agent aims to decide the optimal price on behalf of the buyers by maximizing the utility of the supplier and the broker. However, the independent agents only consider the optimal price with respect to their utility functions. We consider two agents in each approach.

First, Fig. 13 shows that the decentralized approach achieves higher social welfare than the centralized approach, which proves competitiveness of buyers lead to higher social welfare, namely a healthy marketplace. Second, we notice that during training, the decentralized approach converges faster than the centralized one (see the black dashed lines: the left black

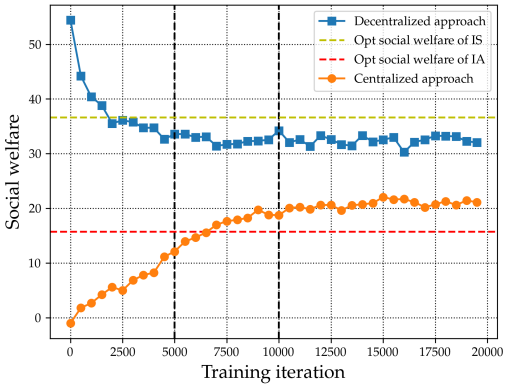


Fig. 13. Centralized approach and decentralized approach with two agents

dashed line is for the decentralized approach and the one on the right is for the centralized approach). We also note that the social welfare for the decentralized approach decreases while the centralized approach increases, but both saturate around 7000 iterations. This echos the fact of conservative behaviour to maximize the payoff in the decentralized approach. However, the centralized approach coordinates the buyers' strategies in favor of the supplier and the broker. Further, this behavior of the centralized approach is in accordance with what was noticed in [11], [12] where it was shown that the centralized approaches fail on relatively simple cooperative multi-agent reinforcement problems, as some states are not explored sufficiently because this leads to a worse team reward in the short term. Finally, both approaches reach social welfare equilibrium in between the ideal scenario (information symmetry) and the worst scenario (information asymmetry). We should emphasize that the Q-learning is essentially using information gained during transactions to estimate the missing information.

VIII. Conclusion

In this paper, we have proposed a mechanism to achieve dynamic service trading in an IIoT market through a DLT. We first analysed an offline approach by considering the problem in both the information symmetric and asymmetric scenarios. Then, an online approach, based on independent learners equipped with Q-learning, has been modeled and solved for marketplaces with dynamic changes. Both approaches show that the joint satisfaction of the supplier, the broker, and the buyer can reach an equilibrium. The important outcome is that the independent learners can approach the joint satisfaction of the information symmetric case even though they are operating within the information asymmetric case. Consequently, this work has shown that the automation of the decision-making of entities in a DLT-assisted IIoT marketplace is possible and that it can be achieved efficiently using the proposed algorithm. Furthermore, through the use of the Byzantine consensus of Hashgraph, the proposed approach is highly efficient compared to DLTs that use proof of work based consensus.

Acknowledgment

This work was supported within the project SerIoT, which has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 780139.

References

[1] T. Riasanow, L. Jäntgen, S. Hermes, M. Böhm, and H. Krcmar, "Core, intertwined, and ecosystem-specific clusters in platform ecosystems: analyzing similarities in the digital transformation of the automotive, blockchain, financial, insurance and IIoT industry," *Electronic Markets*, vol. 31, no. 1, pp. 89–104, 2021. [Online]. Available: https://doi.org/10.1007/s12525-020-00407-6

[2] "Amazon Web Services IoT," (Date last accessed 15-May-2022). [Online]. Available: https://aws.amazon.com/iot/

[3] "Siemens Industrial Edge," (Date last accessed 15-May-2022). [Online]. Available: https://new.siemens.com/uk/en/products/automation/topic-areas/industrial-edge.html

[4] "Online Platforms and the Digital Single Market Opportunities and Challenges for Europe," Communication from the Commision, COM(2016) 288, (Date last accessed 15-May-2022). [Online]. Available: https://ec.europa.eu/transparency/documents-register/detail?ref=COM(2016)288&lang=en

[5] R. Maull, P. Godsiff, C. Mulligan, A. Brown, and B. Kewell, "Distributed ledger technology: Applications and implications," *Strategic Change*, vol. 26, no. 5, pp. 481–489, 2017.

[6] Z. Li, J. Kang, R. Yu, D. Ye, Q. Deng, and Y. Zhang, "Consortium blockchain for secure energy trading in industrial internet of things," *IEEE transactions on industrial informatics*, vol. 14, no. 8, pp. 3690–3700, 2017.

[7] E. Gelenbe, J. Domanska, T. Czàchorski, A. Drosou, and D. Tzovaras, "Security for internet of things: The SerIoT project," in *2018 International Symposium on Networks, Computers and Communications (ISNCC)*. IEEE, 2018, pp. 1–5.

[8] L. Baird, "The swirlds hashgraph consensus algorithm: Fair, fast, byzantine fault tolerance," *Swirlds, Inc. Technical Report SWIRLDS-TR-2016*, vol. 1, 2016.

[9] J. Hu, M. J. Reed, M. Al-Naday, and N. Thomos, "Hybrid blockchain for IoT—energy analysis and reward plan," *Sensors*, vol. 21, no. 1, p. 305, 2021.

[10] P. Bolton, M. Dewatripont *et al.*, *Contract theory*. MIT press, 2005.

[11] T. Rashid, M. Samvelyan, C. Schroeder, G. Farquhar, J. Foerster, and S. Whiteson, "Qmix: Monotonic value function factorisation for deep multi-agent reinforcement learning," in *International Conference on Machine Learning*. PMLR, 2018, pp. 4295–4304.

[12] P. Sunehag, G. Lever, A. Gruslys, W. M. Czarnecki, V. Zambaldi, M. Jaderberg, M. Lanctot, N. Sonnerat, J. Z. Leibo, K. Tuyls *et al.*, "Value-decomposition networks for cooperative multi-agent learning," *arXiv preprint arXiv:1706.05296*, 2017.

[13] A. Asheralieva and D. Niyato, "Distributed dynamic resource management and pricing in the IoT systems with blockchain-as-a-service and uav-enabled mobile edge computing," *IEEE Internet of Things Journal*, vol. 7, no. 3, pp. 1974–1993, 2019.

[14] J. Du, W. Cheng, G. Lu, H. Cao, X. Chu, Z. Zhang, and J. Wang, "Resource pricing and allocation in MEC enabled blockchain systems: An a3c deep reinforcement learning approach," *IEEE Transactions on Network Science and Engineering*, 2021.

[15] H. Yao, T. Mai, J. Wang, Z. Ji, C. Jiang, and Y. Qian, "Resource trading in blockchain-based industrial internet of things," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 6, pp. 3602–3609, 2019.

[16] J. Hu, K. Yang, L. Hu, and K. Wang, "Reward-aided sensing task execution in mobile crowdsensing enabled by energy harvesting," *IEEE Access*, vol. 6, pp. 37 604–37 614, 2018.

[17] K. Liu, X. Qiu, W. Chen, X. Chen, and Z. Zheng, "Optimal pricing mechanism for data market in blockchain-enhanced internet of things," *IEEE Internet of Things Journal*, vol. 6, no. 6, pp. 9748–9761, 2019.

[18] D. Niyato, M. A. Alsheikh, P. Wang, D. I. Kim, and Z. Han, "Market model and optimal pricing scheme of big data and internet of things (IoT)," in *2016 IEEE International Conference on Communications (ICC)*. IEEE, 2016, pp. 1–6.

[19] J. Hu, K. Yang, K. Wang, and K. Zhang, "A blockchain-based reward mechanism for mobile crowdsensing," *IEEE Transactions on Computational Social Systems*, vol. 7, no. 1, pp. 178–191, 2020.

[20] J. Li, T. Liu, D. Niyato, P. Wang, J. Li, and Z. Han, "Contract-theoretic pricing for security deposits in sharded blockchain with internet of things (IoT)," *IEEE Internet of Things Journal*, 2021.

[21] J. Hu, M. Reed, M. Al-Naday, and N. Thomos, "Blockchain-aided flow insertion and verification in software defined networks," in *2020 Global Internet of Things Summit (GIoTS)*. IEEE, 2020, pp. 1–6.

[22] J. Luo, Q. Chen, F. R. Yu, and L. Tang, "Blockchain-enabled software-defined industrial internet of things with deep reinforcement learning," *IEEE Internet of Things Journal*, 2020.

[23] M. Li, F. R. Yu, P. Si, W. Wu, and Y. Zhang, "Resource optimization for delay-tolerant data in blockchain-enabled IoT with edge computing: A deep reinforcement learning approach," *IEEE Internet of Things Journal*, pp. 1–1, 2020.

[24] C. Qiu, X. Wang, H. Yao, J. Du, F. R. Yu, and S. Guo, "Networking integrated cloud-edge-end in IoT: A blockchain-assisted collective q-learning approach," *IEEE Internet of Things Journal*, pp. 1–1, 2020.

[25] I. Psaras, "Decentralised edge-computing and IoT through distributed trust," in *Proceedings of the 16th Annual International Conference on Mobile Systems, Applications, and Services*, 2018, pp. 505–507.

[26] S. Bhattacharyya and F. Lafontaine, "Double-sided moral hazard and the nature of share contracts," *The RAND Journal of Economics*, pp. 761–781, 1995.

[27] J. L. Hernández-Ramos, G. Baldini, R. Neisse, M. Al-Naday, and M. J. Reed, "A policy-based framework in fog enabled internet of things for cooperative ITS," in *2019 Global IoT Summit (GIoTS)*, 2019.

[28] M. Castro, B. Liskov *et al.*, "Practical byzantine fault tolerance," in *OSDI*, vol. 99, no. 1999, 1999, pp. 173–186.

[29] Q. Zhu, S. W. Loke, R. Trujillo-Rasua, F. Jiang, and Y. Xiang, "Applications of distributed ledger technologies to the internet of things: A survey," *ACM Comput. Surv.*, vol. 52, no. 6, nov 2020. [Online]. Available: https://doi.org/10.1145/3359982

[30] M. Saad, J. Spaulding, L. Njilla, C. Kamhoua, S. Shetty, D. Nyang, and D. Mohaisen, "Exploring the attack surface of blockchain: A comprehensive survey," *IEEE Communications Surveys Tutorials*, vol. 22, no. 3, pp. 1977–2008, 2020.

[31] L. Baird, "Hashgraph consensus: fair, fast, byzantine fault tolerance," *Swirlds Tech Report, Tech. Rep.*, 2016.

[32] J. Truby, R. D. Brown, A. Dahdal, and I. Ibrahim, "Blockchain, climate damage, and death: Policy interventions to reduce the carbon emissions, mortality, and net-zero implications of non-fungible tokens and bitcoin," *Energy Research & Social Science*, vol. 88, p. 102499, 2022.

[33] P. Dayan and C. Watkins, "Q-learning," *Machine learning*, vol. 8, no. 3, pp. 279–292, 1992.

[34] R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction*. MIT press, 2018.

[35] E. Nisioti and N. Thomos, "Fast q-learning for improved finite length performance of irregular repetition slotted aloha," *IEEE Transactions on Cognitive Communications and Networking*, vol. 6, no. 2, pp. 844–857, 2019.

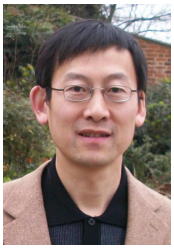
[36] N. Thomos, E. Kurdoglu, P. Frossard, and M. Van der Schaar, "Adaptive prioritized random linear coding and scheduling for layered data delivery from multiple servers," *IEEE Transactions on Multimedia*, vol. 17, no. 6, pp. 893–906, 2015.

Jiejun Hu (M'15) received the Ph.D. from the School of Computer Science and Technology, Jilin University, China, in 2019. She is currently an Assistant Professor at the Lancaster University Leipzig. She was a senior research officer in the University of Essex, UK. Then, she served as a postdoctoral fellow in the Max-Planck Institute for Human Development, Germany. Her research focuses on incentive mechanisms design and game theory in various distributed multi-agent systems, such as the Internet of Things, Mobile CrowdSensing, Software-Defined Networks, blockchain, and digital contact tracing systems.

Martin Reed (M'99) is a full professor in the School of Computer Science and Electronic Engineering at the University of Essex, UK. He has been awarded research funding by UK research councils, Industry and EU research programmes in areas such as network/communication security, IoT security, future Internet architectures, optical network control planes and media transportation over networks, leading to over 100 peer-reviewed papers. His work has resulted in patents, international impact and inclusion in standards by ITU, IETF and 3GPP.

Nikolaos Thomos (S'02,M'06,SM'16) received the Diploma and PhD degrees from the Aristotle University of Thessaloniki, Greece, in 2000 and 2005, respectively. He was a Senior Researcher with the Ecole Polytechnique Federale de Lausanne (EPFL) and the University of Bern, Switzerland. He is currently a Chair Professor in the School of Computer Science and Electronic Engineering at the University of Essex, U.K. His research interests include machine learning for communications, multimedia communications, network coding, semantic communications, information-centric networking, source and channel coding, and signal processing. He is an elected member of the IEEE MMSP Technical Committee (MMSP-TC) for the period 2019-2024. He received the highly esteemed Ambizione Career Award from the Swiss National Science Foundation (SNSF).

Mays F. Al-Naday received the PhD degree from the University of Essex, United Kingdom, in 2015. She is currently a Lecturer in the School of Computer Science and Electronic Engineering, University of Essex, UK. Prior to that, she worked as a senior research officer in the Network Convergence Laboratory (NCL), University of Essex. Her research focuses on future network architectures, including microservice architectures, IoT communications, Fog computing, next generation content delivery networks and security and Quality of Service in 5G and beyond. She has been the organizer of prestigious workshops in Sigcomm 17-18 and IFIP 17. She has actively contributed to a number of EU research projects in the area of future networking architectures.



Kun Yang received his PhD from the Department of Electronic & Electrical Engineering of University College London (UCL), UK. He is currently a Chair Professor in the School of Computer Science & Electronic Engineering, University of Essex, leading the Network Convergence Laboratory (NCL), UK. Before joining in the University of Essex at 2003, he worked at UCL on several European Union (EU) research projects for several years. His main research interests include wireless networks and communications, IoT networking, data and energy integrated networks, mobile edge computing. He manages research projects funded by various sources such as UK EPSRC, EU FP7/H2020 and industries. He has published 150+ journal papers and filed 10 patents. He serves on the editorial boards of both IEEE and non-IEEE journals. He is a Senior Member of IEEE (since 2008) and a Fellow of IET (since 2009).