sustainability

MDPI

Review

Reinforcement Learning in Blockchain-Enabled IIoT Networks: A Survey of Recent Advances and Open Challenges

Furqan Jameel [1,*,†], Uzair Javaid [2,†], Wali Ullah Khan [3,†], Muhammad Naveed Aman [4,†], Haris Pervaiz [5,†] and Riku Jäntti [1,†]

1 Department of Communications and Networking, Aalto University, 02150 Espoo, Finland; riku.jantti@aalto.fi
2 Department of Electrical and Computer Engineering, National University of Singapore, 4 Engineering Drive 3, Singapore 117583, Singapore; uzair.javaid@u.nus.edu
3 School of Information Science and Engineering, Shandong University, Qingdao 266237, China; waliullahkhan30@gmail.com
4 School of Computing, National University of Singapore, 13 Computing Drive, Singapore 117417, Singapore; naveed@comp.nus.edu.sg
5 School of Computing and Communications, Lancaster University, Lancaster LA1 4WA, UK; h.b.pervaiz@lancaster.ac.uk
* Correspondence: furqanjameel01@gmail.com
† These authors contributed equally to this work.

Received: 27 May 2020; Accepted: 23 June 2020; Published: 24 June 2020

Abstract: Blockchain is emerging as a promising candidate for the uberization of Internet services. It is a decentralized, secure, and auditable solution for exchanging, and authenticating information via transactions, without the need of a trusted third party. Therefore, blockchain technology has recently been integrated with industrial Internet-of-things (IIoT) networks to help realize the fourth industrial revolution, Industry 4.0. Though blockchain-enabled IIoT networks may have the potential to support the services and demands of next-generation networks, the gap analysis presented in this work highlights some of the areas that need improvement. Based on these observations, the article then promotes the utility of reinforcement learning (RL) techniques to address some of the major issues of blockchain-enabled IIoT networks such as block time minimization and transaction throughput enhancement. This is followed by a comprehensive case study where a Q-learning technique is used for minimizing the occurrence of forking events by reducing the transmission delays for a miner. Extensive simulations have been performed and the results have been obtained for the average transmission delay which relates to the forking events. The obtained results demonstrate that the Q-learning approach outperforms the greedy policy while having a reasonable level of complexity. To further develop the blockchain-enabled IIoT networks, some future research directions are also documented. While this article highlights the applications of RL techniques in blockchain-enabled IIoT networks, the provided insights and results could pave the way for rapid adoption of blockchain technology.

Keywords: blockchain; Industrial Internet-of-things (IIoT); Industry 4.0; Q-learning; reinforcement learning (RL)

1. Introduction

Blockchain is steadily gaining traction due to its applications in banking, supply chain management, and cybersecurity [1–3]. Besides simplifying the business processes and reducing errors

in verification, its integration with Internet-of-things (IoT) has recently been discussed extensively among the research community. As one of the key enablers of Industry 4.0 and an emerging offshoot of IoT, the Industrial IoT (IIoT) networks are paving their way in various commercial and social sectors such as retailing, manufacturing, logistics, pervasive monitoring, security surveillance, healthcare, and home automation [4–7]. Moreover, with the recent developments in wireless communications and sensor network technologies, an increasing number of devices are being introduced in the IIoT space, where raw data are locally captured and processed to support decision-based processes. These devices can communicate and interact with each other as well as share and process information independent of human intervention [8]. Therefore, they must be made secure to preserve data integrity as well as ensure resource availability and computing reliability.

Blockchain technology introduces a new design paradigm for next-generation transaction-based applications that, together with a distributed and shared public/private ledger and a collective consensus mechanism, build trust, transparency, and accountability in a system. A summary of these processes, illustrating a transaction from A to B, is given in Figure 1. Due to the transparent nature of blockchain technology, it is highlighted by the academia and the industry alike, as a potential solution for efficiently managing massive IIoT networks [5,6,9]. Moreover, decentralized IIoT devices and trustless network architectures are expected to play a key role in the advancement of IIoT networks where data will be processed locally at the site of generation and not in a centralized manner. It will also facilitate device connectivity and make data storage trustless through devices and sensors that can operate without relying on a central authority. A blockchain can also offer IIoT devices a secure infrastructure that is robust against single-point-of-failure [10,11]. Because decentralized networks have multiple entry points, they adhere to resilience and fault tolerance of the networks. In addition to this, IIoT infrastructures can become more accessible by using distributed ledger technologies. A comparison of blockchain with different ledger technologies is given in Table 1. Although there is a need to perform a detailed cost analysis of private blockchain networks, certain public blockchains platforms (e.g., IOTA) have zero transaction fees. Thus, the associated costs for operating centralized IIoT systems can also be reduced significantly by decentralizing the IIoT networks using blockchain.

Since blockchains do not rely on intermediaries for their operation, they can automate services through code, and by distributing control in the network via consensus algorithms. In such networks, trust between different IIoT devices is established through distributed consensus protocols, thereby removing the need of a trusted centralized service provider. This insight is also very important from the perspective of self-sustainability in IIoT networks [12]. The notion of "distributed autonomous corporations" can be implemented on decentralized IIoT systems that can operate independently according to a pre-defined logic. Moreover, smart contracts (a set of encoded logic that can be used to create agreements when a certain set of conditions are met) can also be used for verifying a function or data and automating a procedure. This capability can be useful for many IIoT applications. For instance, a user/device can authorize a certain payment when a set of conditions indicate that the delivery of a product/service has been completed. This way interactions between user-user, user-device, or device-device can be handled transparently.

Furthermore, the systematic and, in particular, automatic processing of information or data with help from computers is the core task of computer science. For such processing, computer programs are developed, which are traditionally engineered by people. Each program can have one or a set of problems to address. The nature of these problems can range from simple to extremely complex and time-consuming or even computationally infeasible to solve. This can directly affect the processing of the input information and the creation of a program. Thus, addressing these problems and meeting optimal program design requirements remain key concerns in computer science. One of the most debated solutions to address these concerns is the so-called artificial intelligence (AI) that has the potential to define better program design principles [13].

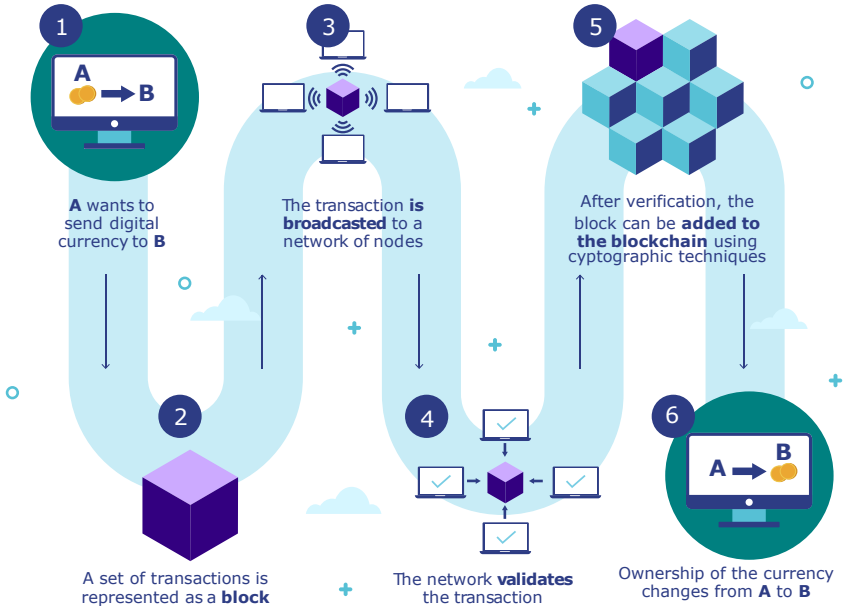


Figure 1. An illustration of exchanging digital currency from A to B in a blockchain.

The basic goal of AI is to solve problems that are not easy for humans, for which a certain form of intelligence seems necessary. For example, the human decision-making process can be supported by expert systems. Rule-based systems are very common in this environment. The manual modeling of the facts and rules required for these systems as a knowledge base is very complex, expensive, and often also prone to errors. This difficulty is widely known as the knowledge acquisition bottleneck. Can the knowledge not be derived automatically from experiences that are available in the form of data? The initial development of machine learning as a part of AI can be interpreted as the answer to solving this problem. In the meantime, machine learning can even be seen as a key technology for solving AI tasks and is also used in numerous application areas that are not directly assigned to AI.

1.1. Characteristics and Overview of Reinforcement Learning Techniques

Reinforcement learning (RL) is a sub-branch of AI or, more generally, of computer science that deals with having computers solve problems without explicitly programming them. It is a sequential learning process where agents automatically adjust their policies by observing the result of their rewards. More specifically, the RL techniques can be based on the Markov decision process (MDP) that consists of an environment and a set of agents [14,15]. The environment transitions into a new state once the agent takes some action and receives a reward in response. Due to state transitions, the dynamics of the environment can be modeled as a sequential decision-making process. Interaction with the surrounding world can be considered a foundation of human learning. Machine learning through reinforcement follows this basic paradigm learning. At its center, an agent tries to achieve a long-term goal by practicing meaningful actions. It can then provide feedback on these actions to gradually improve output. This approach received special attention in 2015 and 2016 when the best human players of the board game Go went against AlphaGo. Compared to chess, Go offers significantly more combinations and it is much more difficult to assess which player is currently in the lead. Moreover, this learning approach has also recently seen many novel applications in a wide range of domains.

In contrast to supervised and semi-supervised approaches, RL techniques generally do not need labeled data or prior information regarding the environment. This characteristic of RL makes it suitable for blockchain technology. The RL techniques can be broadly categorized into two types, i.e., model-based RL and model-free RL [16]. A brief description of these two types is given below:

1.1.1. Model-Based RL

The model-based approaches assume that the agent has access to the model of an environment. The model of an environment can be a function that predicts the transition probabilities and stat-function. Under these circumstances, the agent can have a better understanding of the environment as it can plan and think ahead about several possible choices. This approach generally improves the efficiency of the agent and help by planning the policy. A good example of this approach is AlphaZero [17].

In recent years, an extensive number of methods have been proposed for model-based RL. One of the emerging techniques is model-predictive control for selecting actions of the agent [18]. In this way, the agent formulates an optimal plan (i.e., actions taken over a long period of time) after observing the state of the environment. The learning agent prepares a new plan after each new interaction with the environment. Another popular approach for model-based RL is data augmentation [19]. It uses a learning algorithm to train the agent and either use fictitious experiences or augments the real experience with a fictitious one. Another approach, called embedded planning, makes use of subroutine which acts as side information. This provides the agent with the capability to choose a particular plan and ignore the other plans based that do not provide optimal policy.

1.1.2. Model-Free RL

Although there are many advantages of model-based RL, yet it is difficult to train the model-based RL. Furthermore, it is often difficult to find the ground-truth model for training the agent. The bias in the model could also be exploited by agents, thereby, performing sub-optimally in a practical setting. In these conditions, the best approach is to adopt a model-free RL approach to learn different aspects of the environment. This generally leads to a low bias training and is considered more feasible when multiple factors are affecting the reward. A brief discussion of these models is provided as follows:

- **Q-learning**: Q-learning is the most studied and well-known RL technique that has been used for solving different sorts of problems. In Q-learning, an agent takes action based on the Q-values in the Q-table [20]. There are four major components of a Q-network, i.e., a stage set, an action set, a reward, and transition probabilities. For each state, the agent executes some action under its pre-defined policy. Subsequently, the agent adopts the policy such that it has maximum Q-value. After each sequence, the agent revises the Q-table for a more accurate estimation of the Q-values and updating the policy of the agent. Thus, in due course of time and after many steps, the policy converges to the optimal policy of the agent.
- **Multi-armed Bandit Learning**: The agent in the multi-armed bandit approach selects the action without having the state information of the environment. Relevant to the action performed by the action time step, the agent receives a reward that is maximized in the next iterations [14]. Since the agent is unaware of the environment and the associated reward, there always exists a tradeoff between exploitation and exploration. Due to this reason, the multi-armed bandit learning, although lower in complexity, is not efficient for highly dynamic environments.
- **Actor-critic Learning**: The actor-critic learning divides the agent into two roles, i.e., actor and critic [20]. The action selection policies over the action space are represented by the actor. In contrast, the critic is the observer which anticipates the expected reward received in the future by passing through the same state. Based on the observed reward, the state values are updated by the critic. Here, the critic can be considered as a trainer which trains the actor to improve its stability and select the optimal action in each state. Using the probability density function given by the actor, the size of the action space does not increase the complexity of the action selection. Therefore, it is much suitable for continuous and large action state spaces.
- **Miscellaneous**: Other techniques include policy optimization, soft actor-critic, and deep policy optimization techniques. These techniques are mostly variants of the above-mentioned RL techniques that use entropy regularization, and stochastic policies to stabilize RL. Sometimes

the conventional RL techniques may not be feasible for large-scale networks. In other words, the dimensionality of state, action, and reward could render the problem difficult to solve. Thus, for high dimensional optimization, deep neural networks have been proposed to work along with conventional RL techniques to improve the network. These deep neural networks can be used to estimate the Q-values during each iteration, also known as deep Q-network [20]. The agent can select the maximum estimated value and store the experience again to train the neural network for estimation. Nevertheless, it is much more complex and computationally exhaustive than conventional RL techniques.

1.2. Motivation

There exist many issues and caveats in different avenues of blockchain-enabled IIoT networks that remain unresolved and to be addressed. These issues restrict the integration of blockchain with IIoT as they have different system limiting factors or trade-offs under different circumstances. This trade-off is similar in nature to CAP (Consistency, Availability, tolerance to network Partitions) theorem of distributed systems that states: "A robust and distributed system can only simultaneously provide two out of its three properties" [21]. Similarly, the integration of blockchain can also be considered as a trade-off challenge. Therefore, this article first presents a review of current literature to highlight the challenges associated with blockchain-IIoT networks and then, it discusses how they can be addressed.

1.3. Related Surveys and Our Contributions

To address the concerns discussed above, this article formulates a gap analysis of recent literature, which is used to reflect and highlight some of the overlooked aspects of blockchain-enabled IIoT networks. Unlike RL surveys published in other articles, we present our gap analysis through an abstract representation of a blockchain-IIoT network. We believe that such representation has not yet been discussed in the literature. Subsequently, similar to other surveys, our article advocates the need for employing RL techniques to optimize their performance. In particular, it discusses some potential applications of RL in blockchain-enabled IIoT networks ranging from minimizing block time to improving transaction throughput. On the other hand, unlike the surveys of recent literature, our article also provides a case study where a Q-learning technique is used to minimize forking events. The simulation results of the RL technique demonstrate improvements when compared with a conventional greedy policy. Finally, some future research directions are provided for researchers working in academia and industry. Thus, the main contributions made in this paper unlike other surveys, are the following:

1. Identifying the integration challenges of blockchain with IIoT and formulating a problem statement for a case study.
2. A thorough review of the current paradigms in blockchain and their associated consensus algorithms.
3. A review of reinforcement learning, its characteristics, and how they can help address the integration problems of blockchain and IIoT.
4. An abstract representation of a blockchain-IIoT network with three-layer network architecture, i.e., Physical Layer, Network Layer, and Application Layer.
5. A concise review of how RL can be used to power different avenues in blockchain-enabled IIoT networks.
6. A case study to demonstrate the feasibility and improvements introduced by using RL in a blockchain-IIoT network.

Table 1. Comparison of distributed ledger technologies.

Attributes	Blockchain	Tangle	Hashgraph	Sidechain
Data Structure	Linked list	Directed Acyclic Graph (DAG)	DAG	Linked lists
Topologies	Public	Private	Private	Semi-public
Consensus	Proof-of- Work-SHA256 Proof-of- Stake	Proof-of-Work-Hashcash	Virtual Voting	Proof-of-Work-Ethash
Transaction Structure	Blocks	Sites	Gossip Event-based	Consortium Chain
Transaction Fee	Yes	No	No	Yes
Transactions Per second	4–12	500–800	>200,000	3000–20,000
Verification Time	Order(minutes)	Order(seconds)	Order(seconds)	Order(minutes)
Privacy	Moderate	Low	Low	Moderate
Security	Very High	High	High	High
Maturity	Many Implementations	Experimental Phase	Experimental Phase	Experimental Phase
Platforms	Bitcoin, Ethereum	IOTA	Hyperledger	Ethereum, Monax, Elements
Copyright	Open Source	Open Source	Patented	Open Source

1.4. Organization

The remainder of the paper is organized as follows. In Section 2, preliminaries of blockchain-based IIoT networks is provided. In Section 3, the existing gap analysis in the literature is outlined. This is followed by Section 4 which highlights the applications of RL in blockchain-based IIoT networks. Next, a case study is then presented in Section 5 for the minimization of forking. Finally, Section 6 provides conclusions and potential future research directions.

2. Blockchain-enabled IIoT Networks

Industry 4.0 represents the fourth industrial revolution that will facilitate IIoT with adaptive and autonomous systems that can self-heal and self-learn. IIoT aims to promote multi-disciplinary businesses and industries by realizing intelligent industrialization [22–25]. It enables smarter industrial processes by incorporating AI with big data technologies for exploiting the massively produced and communicated data. The recent surge in data volumes generated by IIoT environments and then sent to centralized servers, present some security concerns like a single point of failure, data integrity, and in particular, scalability. Decentralized IIoT network architectures are expected to address these issues and play a key role in the advancement of IIoT, where data will be locally processed at the site of generation and not in a centralized manner. Over the years, blockchain technology has moved from the phase of inception to rapid research and development, as shown in Figure 2. Thus, it is high time to explore the applications of this technology in IIoT networks to solve the security problems and provide decentralized IIoT solutions via transparent, immutable, and distributed system design principles. Moreover, decentralizing IIoT can provide the following benefits:

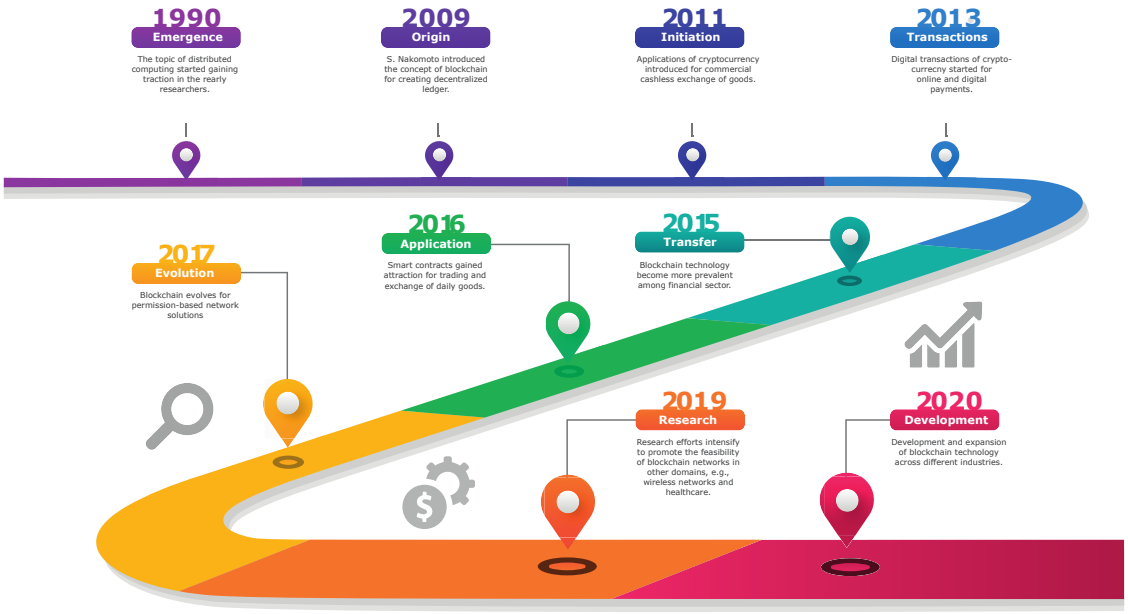


Figure 2. Timeline of blockchain technology over the years since its inception.

2.1. Improved Security

A blockchain can offer IIoT devices a security infrastructure that is robust against a single-point-of-failure. A special feature of blockchain technology is the decentralized structure of the network. In a centralized network, transactions between network participants are always made with the help of a central instance (node). This intermediary can control all movements within the network. The individual members communicate with each other via this central point and not directly with each other.

A decentralized structure dispenses with such an instance so that direct communication with one another is possible. Such a network cannot be controlled from the outside. The users of the network can be distributed worldwide, but still have a uniform, synchronized database. The most common term for this is the "peer-to-peer network". Decentralized networks have multiple entry points that provide resilience and fault tolerance to networks. Moreover, public key infrastructure makes attacking a blockchain incredibly difficult because any data originating from anywhere other than the origin, i.e., genesis block, will be null and useless in the network.

2.2. Data Integrity

A blockchain uses cryptography for referencing its blocks, i.e, it adds a hash value of previous blocks in its succeeding blocks. Despite the low residual risk of hacking, blockchain technology ensures a high level of security, as the data is distributed locally, accessible to all users, and encrypted. It offers a high degree of reliability since the data is saved redundantly on all full nodes. By dispensing with intermediaries such as banks, it is possible to carry out the transactions directly with one another. This allows faster processing and, particularly in regions with a less developed legal system, enables contracts or transfers to be executed correctly and securely. This way, a block, and its data can be verified completely by checking the reference hash in a block. This property of blockchain preserves data integrity and makes it extremely difficult to tamper with data.

2.3. Cost Effectiveness

IIoT infrastructures can become more affordable when security vulnerabilities in their systems are removed by decentralizing the networks and store their data in distributed ledgers. Mutual trust in the parties involved in a transaction is not necessary with blockchain networks; rather, the technology behind it ensures secure transaction processing [26]. Furthermore, transparency in a blockchain network is extremely important. In this way, the entire transaction history is clearly shown and users can view it at any time. Blockchain networks work independently and autonomously, which is why external influences do not affect the network. In traditional IIoT systems, the service providers usually have a monopoly on their operation and the cost of supporting devices [27]. By using distributed ledger technologies, service providers and their monopolies can potentially be completely removed, thereby making IIoT more accessible. Moreover, the associated costs for operating centralized IIoT systems can also be eliminated.

2.4. Trustless

A blockchain does not rely on intermediaries for its operation and therefore, can automate services through code and by distributing control in the network [28]. For the participants in a blockchain network, the private key is generated in the form of a random number and the public key is derived from it. The user's address is generated from the public key as an alphanumeric value. This is also called "pay to public key hash". The public key of a user can be recognized at any time by the other network participants, in contrast, the private key is secret and is used for the decryption and signature of the transactions. If two users want to execute a transaction within the blockchain network, the sender encrypts it with the recipient's public key. Decryption, i.e., making it readable again, is only possible if the recipient decrypts the transaction with his private key. To prove that the transaction came from the sender and not from an unauthorized third party, the sender signs it with his private key. The recipient can now use the sender's public key to ensure that the transaction originated from the sender [29]. To increase the security of individual transactions so that, for example, the key cannot be used to conclude other transactions of a participant, there is the possibility that network participants use a new key pair for each transaction. Trust between users and devices in an IIoT system can then be established by using distributed ledgers with smart contracts, which will eliminate the need to place trust in centralized service providers.

2.5. Autonomy

The notion of "distributed autonomous corporations" can be implemented on a decentralized IIoT system that can operate independently according to some predefined logic. There is no central body that authenticates the participants, which is why all users have the same legitimation [30]. The nodes of a blockchain network store and manage the entire transaction history of a blockchain in an unchangeable form. Over time, however, this transaction history takes up an enormous amount of storage capacity [31]. In contrast, full nodes store the entire blockchain database. This can potentially remove intermediaries and central authorities to facilitate automation in IIoT systems.

2.6. Smart Contracts

These are computer programs, i.e., coded logic, that can be used to create agreements that are executed when a certain condition or a set of conditions are met [32]. Practical examples of such a blockchain network are Ethereum or Bitcoin. Etherum offers a cryptocurrency called "Ether" which is a smart contract platform. There are a variety of different blockchain applications that are not limited to cryptocurrency since users can implement their functions [33]. They can also be used for verifying a function as well as data, and to automate a process. This can be very useful for many applications in IIoT. For instance, a user/device can authorize payment when a set of conditions

indicate that the delivery of a product/service has been completed. This way interactions between user-user, device-device, or user-device can be handled transparently.

A blockchain is an online, digital ledger that is globally distributed [34]. The nature of this ledger can be public, private, or semi-private. Note that public ledgers are permissionless while private ones are permissioned. It defines a distributed system design paradigm that uses cryptography, a public key infrastructure (PKI), and economic modeling. This integration of primitives is applied together to a peer-to-peer (p2p) network as well as a shared consensus algorithm. The consensus is used to achieve synchronization among the distributed ledger and it typically operates on a huge number of nodes and/or devices [35]. Moreover, on a blockchain, any kind of information and anything of value can be stored such as digital assets, deeds, identities, and even votes can be securely stored, moved, and managed. A brief description of the different components of blockchain is provided in Figure 3. The set of its properties of decentralization, immutability, transparency, and fault-tolerance render is suitable for decentralized IIoT environments.

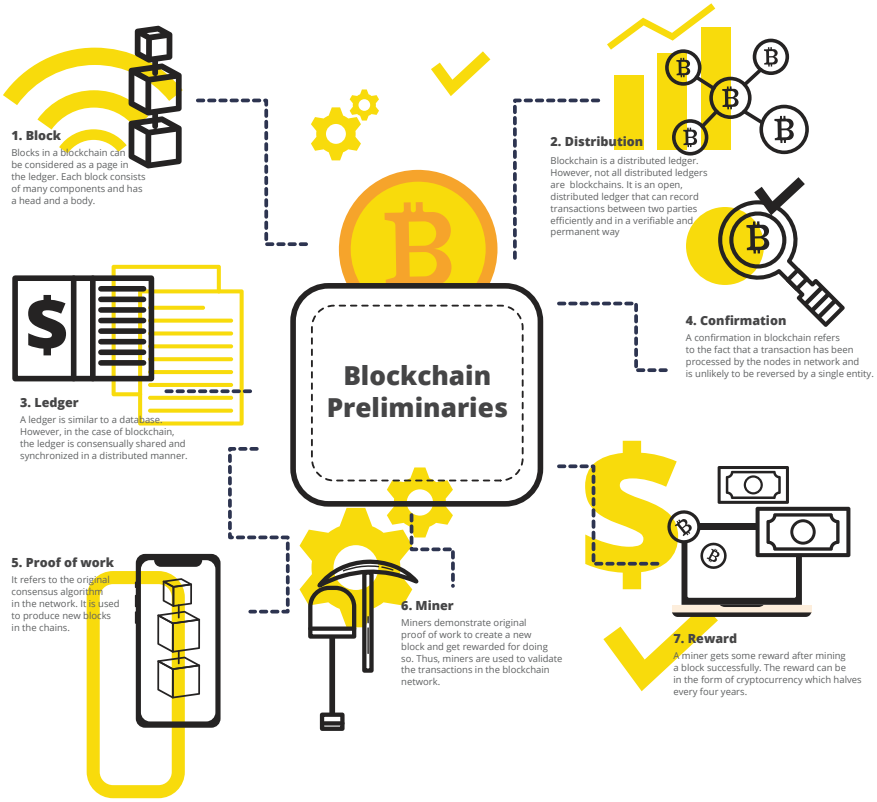


Figure 3. A brief illustration of blockchain primitives including ledger, miner and block.

It can be inferred that the main constituent in a blockchain is 'blocks'. A block is a set of transactions, i.e., it collates tuples of transactions together concerning a specific period. At any time t in a blockchain network, the users generate hundreds of transactions. These transactions are initially unconfirmed and need to be verified so that they be eligible to be written in the blockchain. This requires the confirmation of miners, which add unconfirmed transactions together and form a block. Note that miners are blockchain entities that are responsible for generating new blocks. Thus, the first block in a blockchain is called the 'genesis' block and all the blocks added after it is called the 'successor' blocks. The successors are added to the genesis block in chronological order, i.e., the genesis block b_g is hashed and stored in the second block b_{g+1}. The hash of b_{g+1} is stored into b_{g+2} and so on. This way, each block has a hash pointer of the previous block and to change the content of one

block requires changing all of its preceding blocks until genesis. This forms the basis of a blockchain and can be formulated as:

$$\mathcal{B} = (b_g, b_{g+1}, b_{g+2}, \cdots) \tag{1}$$

Moreover, as mentioned before, a block contains a set of transactions. It can also contain application relevant data as well as certain meta-data such as block header of the previous block. Note that a block header is a common term used for representing the hash of the previous block. Thus, we can formulate a block in the following way:

$$\mathcal{B} = (b_g, b_{g+1}, b_{g+2}, \cdots) \implies b_g = \sum_{i=0}^{n} (tx(\cdot), data, metadata)_{gi} \tag{2}$$

where $\mathcal{B}$ represents a blockchain, b_g the chronologically appended blocks in it, and $tx(\cdot)$ represents the set of transactions included in a block. From Equation (2), we can see that a block can contain n instances of transactions, data, and other information depending upon the nature of the application it is designed for.

A transaction set $tx(\cdot)$ is a set of instructions that changes the ownership of digital assets from one user to another. Note that a digital asset can be a virtually valued currency, a token, or simply a deed, etc. The ownership of these assets is changed via the PKI framework offered by a blockchain. Mathematically, we can formulate a transaction in the following way:

$$tx = t_{in}Num \parallel t_{in} \parallel t_{out}Num \parallel t_{out} \parallel nonce \parallel data \parallel t \tag{3}$$

where t_{in} and t_{out} are input, output vectors of a transaction that represent both the sender and recipient. $t_{in}Num$ and $t_{out}Num$ are the number of transactions relative to a timestamp t, *nonce* represents the challenge for miners to mine the transaction and validate it, whereas *data* field is an extra field that can store transaction relevant information.

Furthermore, in addition to storing block headers in metadata, a blockchain also stores the proof of a block in it. The proof is generated by the miners, which is a unique and deterministic function. Typical blockchain applications use the proof-of-work (PoW) consensus algorithm which can be generally formulated as:

$$\mathtt{pr}(p \leq t) = p/2^{256}, \ \forall \ 256 - bit\ based\ systems \tag{4}$$

where the probability of finding a proof p for a target t is $\mathtt{pr}(p \leq t)$. When a miner finds the target value, it broadcasts its proof in the network. The other miners then verify this proof and validate the block. Once the block is validated and accepted by the majority of the network, it is then finally added to the blockchain. Moreover, the time taken by a miner to find proof for a block can be given as:

$$t_p = 2^{256}/hashRate \tag{5}$$

where t_p is the time required to find a proof for a block and *hashRate* represents the number of hashes a miner can generate per second. Generally, hundreds of thousands of hashes are required per second to mine a block in a feasible time frame. Therefore, PoW based systems arguable require high computational resources and complexity.

We can now deduct that blockchain technology introduces a whole new design paradigm for next-generation applications that, together with a distributed and shared public/private ledger and a collective consensus mechanism, build trust, transparency, and accountability in a system. Thus, many different avenues have been already identified to benefit from its adoption such as financial institutions, smart power grids and cities, supply chain management frameworks, and cyber-physical systems [36–38]. Besides simplifying business processes and providing transparency in system operations, its integration with IIoT has recently been greatly discussed among the research community.

As one of the key enablers of I4 and an emerging offshoot of IoT, IIoT networks are paving their way in various commercial, and social sectors such as retailing, manufacturing, logistics, pervasive monitoring, security surveillance, healthcare, transportation systems, and home automation, etc. [34,39–41]. Moreover, with the recent developments in wireless communications and sensor network technologies, an increasing number of devices are being introduced in the IIoT space, where raw data are locally captured and processed to support decision-based processes. These devices can communicate and interact with each other as well as share and process information independent of human intervention. Therefore, they must be made secure to ensure data integrity as well as resource availability and computing reliability [42].

3. Recent Studies and Gap Analysis

This section discusses some of the recent works done in blockchain-based IIoT networks while also providing a gap analysis. Although there is little consensus on the number of layers in a blockchain (Generally, a different arrangement of six layers (i.e., Data layer, Network layer, Consensus layer, Incentive layer, Contract layer, and Application layer) can be found in blockchain literature [43]. However, due to the lack of overlap between blockchain and IIoT layered architecture, it is not straightforward to deduce the optimal number of layers. Therefore, we resort to three layers that may incorporate operations of other sub-layers.), there are generally three layers [44] (i.e., Physical (Perception), Network, and Application) having common functionality in IIoT and blockchain technology.

3.1. Physical Layer

For the physical layer, blockchain helps to facilitate key management in IIoT devices through their gateways that can act as the agent of a cluster of IIoT devices [4]. In this regard, the authors in [45,46] address the issue of digital provenance in IoT based environments and vehicular networks, and propose a hardware primitive based framework that uses physical unclonable functions (PUFs), blockchain, and smart contracts. A PUF can be defined as a system that maps a set of challenges to a set of responses based on the physical microstructure of a device. In [47], Jiang et al. analyzed the wireless power transmission aspect of blockchain networks. The authors in [48] present a chip-level blockchain-based solution for the IIoT networks. They propose the integration of IoT with blockchain that uses a physical address inside a semiconductor chip mounted into an IIoT device. Furthermore, the authors in [49] present a new dimension to PUFs and introduce optical PUFs as a physical trust root for blockchain-based applications. They stress that optical PUFs (o-PUFs) can be successfully used as random number generators to generate bit strings. They explain how the bit strings can be generated by using a coherent light source that can be varied to create different PUF challenge-response pairs. The generated bit-strings can then be used as symmetric keys for asymmetric encryption.

3.2. Network Layer

For the network layer, there are studies that focus on the merger of blockchain with other technologies like software-defined networks and edge computing [50,51]. The authors in [52] present four major blockchain trends. Note that we differentiate among the blockchain technology trends concerning their consensus protocols and not their applications. Blockchain 1.0 (e.g., Bitcoin, Ethereum) represents the typical blockchain data structure that starts with a genesis block and adds succeeding blocks to it in a chronological manner. These are the blockchain paradigms whose consensus protocols are proof-of-work. To exploit these protocols for system designs, one example includes [53] in which the authors propose a blockchain-based payment scheme for remote villages with intermittent Internet connectivity. Blockchain 1.0 however, has shown weak scalability and favored mining, thereby highlighting the need for a new, more effective protocol. A more advanced version is therefore introduced as Blockchain 2.0 (e.g., Ethereum 2.0) that uses smart contracts and proof-of-stake protocols to protect the network [54]. It has also resulted in a significant reduction of the cost of verification and

allows a transparent contract definition to prevent fraud and hacking. Blockchain 3.0 is another advancement over the previous version and uses DAGs in which there are no blocks but sites, i.e., transactions committed by IIoT devices. Each device represents a transaction and the connections (direct edges) between transactions represent their validation. The authors of [55] exploit Blockchain 1.0, 2.0, and 3.0 to propose deterministic cross-blockchain token transfers (DeXTT), a cross-blockchain transfer protocol that can be used to transfer tokens, i.e., digital assets from one blockchain to another. The most recent version of the blockchain is Blockchain 4.0 which is built to increase the degree of trust and privacy in Industry 4.0. Blockchain 4.0 is an extension of Blockchain 3.0 to make it feasible for real-life business scenarios. The consensus protocol for this trend includes virtual voting, blockchain consortium (federated blockchains), and gossip-based protocol variants. IIoT data collection, approval of workflows, asset and supply chain management are some of the key business processes that could be enabled by Blockchain 4.0.

3.3. Application Layer

In blockchain-enabled IIoT networks, the application layer deals with the interfacing and controlling issues [56]. This layer can also be used to develop programmable currency, smart contracts, and REST APIs. For instance, the authors of [57] studied the reserve price of advertisement inventory in blockchain networks and analyzed its impact on the online market. Subsequently, they formulated an auction model for multi-channel sales of the advertisement inventory. Qian et al. in [58] highlighted the utility and importance of the application layer in the healthcare domain, similar to the work of [59]. They argued that there is a need for secure application layer architecture for protecting user privacy and storing encrypted data in the cloud or edge. Another use case of blockchain is for mitigating IIoT device based distributed denial of service (DDoS) attacks [2]. The authors used Ethereum with smart contracts that verify IIoT devices and enforces a bandwidth limit on them beyond which they cannot operate.

4. Applications of RL in Blockchain-Enabled IIoT Networks

RL techniques are fundamentally different from the other featured learning, where at the beginning of the learning process, the training data are available, be in accommodation with or without a label as examples or observations. That is why RL is also called learning through interaction. While in supervised learning, knowing the correct target values for the training data can determine exactly whether a decision is right or wrong (classification) or how far a prediction is from the correct value (regression), the agent in RL does not usually learn whether a decision is the best. Since the wealth of experience only grows with its interaction, the agent often only knows for a part of the possible actions that have led to a greater or smaller reward in the past. To solve this problem, typically previously unknown (not evaluated) actions are performed based on trial and error to increase the scope of activities within the scope of his strategy.

Aforementioned in view, RL is the least specified in the blockchain domain, so this learning form is ideal for scenarios in which the provision of training data is difficult and an agent has to make a strategy for a series of decisions. This interactive learning makes the RL techniques suitable for blockchain-enabled IIoT networks. This section builds on this thesis and provides details of how different RL techniques can be applied to IIoT-blockchain networks to improve their performance. A summary of these applications is provided in Figure 4.

4.1. Minimizing Forking Events

In general, forking is an event in the blockchain that takes place when a blockchain diverges into two potential chains. Since the miners in the blockchain need to use common consensus algorithms to maintain the history of blockchain, a forking event indicates a scenario where the miners clash and have a conflict of consensus agreement. This may result in creating forks that are both short and long. From the perspective of Blockchain 1.0, the forking may be caused when the nodes having

completed proof-of-work, do not convey the results to other computing nodes. To avoid the forking events, the transmission delay at the miner can be reduced with the help of RL. The agents can be trained for an IIoT environment that develops an optimal policy for minimizing such delays.

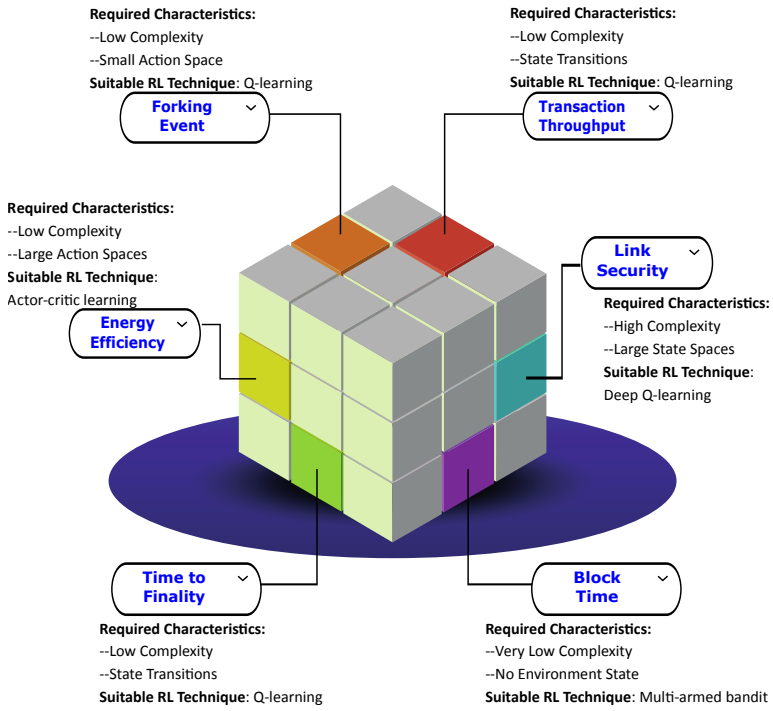


Figure 4. Applications of RL techniques for blockchain-enabled IIoT networks. The Q-learning technique appears to be more suitable for improving transaction throughput and minimizing forking events. Moreover, energy efficiency and link security can be improved using actor-critic learning and deep Q-learning techniques, respectively. Finally, time to finality and the block time is expected to be reduced, respectively, with the help of Q-learning and multi-armed bandit learning.

4.2. Improving Energy Efficiency

Energy is an important aspect of providing communication infrastructure to IIoT-blockchain networks [3,60]. Due to this reason, the importance of energy-efficient communication techniques cannot be overstated. The energy-constrained miner can become the point of failure in case the energy resources are not utilized efficiently. Besides this, the forking events may also result in re-computation of the proof-of-work at the miner side, thereby, reducing the energy budget. These situations may entail an extra cost in terms of consuming unnecessary energy of devices [61]. To avoid this, RL techniques can be very helpful when applied correctly. We anticipate that these types of networks can be optimized using actor-critic learning. Due to the continuous and dynamic nature of the energy consumption of blockchain-enabled IIoT devices, it is important that the versatile actor-critic learning model may be employed.

4.3. Time to Finality Minimization

The term finality refers to the confirmation message that once committed to the blockchain, the well-formed blocks would not be tampered with. This is important since the blockchain users need to ensure that the transactions cannot be reversed or arbitrarily changed once a transaction goes through [31]. Although the consensus protocols have been designed to reach finality in a smaller time, their impact on IIoT devices is not clear yet. Therefore, it is important to reduce the probabilistic time to finality for proof-of-work protocols. In this regard, it is important to avoid selfish mining by devices. An RL system can be used to detect selfish miners that consume the resources unnecessarily.

For instance, the Q-learning technique can be used to minimize the probabilistic time to the finality of proof-of-work for blockchain-enabled IIoT networks. The agents can update themselves by learning from the received rewards and probabilistically improving the time to finality [62].

4.4. Enhancing Transaction Throughput

Transaction throughput refers to the number of transactions per second [63]. The current number of transactions in IIoT-blockchain networks suffers from scalability problems. One convenient solution can simply be to increase the number of transactions per block. The other way can be to increase the frequency with which the blocks are added into the network. Thus, recording more transactions could have a reverse effect on the decentralization of blockchain-enabled IIoT networks. It may also increase the mining time since a miner needs to check the validity of all signatures on the transactions before mining a block. Because of such intricacies, the optimal block size and frequency of adding new blocks are highly application- and resource-dependent while the RL techniques can be used to efficiently regulate the network dynamics [64,65]. Moreover, the optimal policies and the tradeoffs between the transaction throughput and decentralization can also be identified with the help of RL techniques [66]. The agent can also provide the specific set of actions needed to increase the throughput while not compromising the decentralization of IIoT-blockchain networks.

4.5. Improving Link Security

Another important aspect of IIoT-blockchain networks is linked to security [67]. This becomes critical when miners are sharing sensitive information or exchanging acknowledgment messages. Due to the broadcast nature of messages exchanged between IIoT-blockchain devices, it is important to secure the links through physical layer security (PLS) techniques [8]. The PLS exploits the randomness of a wireless channel to confuse the eavesdroppers. On other occasions, artificial noise can also be added by a friendly jammer to protect the communication channel. The RL can be extensively used for the provisioning of link security for IIoT-blockchain networks. Multi-armed bandit techniques can be used to identify the nearby eavesdroppers in the blockchain network. Deep Q-learning can be applied to introduce the artificial noise in the network, without damaging the quality of the legitimate link. Due to the dynamicity of these learning techniques, they can be applied easily in many indoor and outdoor link security cases [68].

4.6. Average Block Time Reduction

Block time, or block interval, can be defined as the total amount of it takes to mine a block. Due to the high variability in time for mining a block, the average block time is more preferred in large-scale networks. Therefore, the average block time of a blockchain-enabled IIoT network is directly related to the complexity of the proof-of-work algorithm. It may be noted that some existing platforms (e.g., Ethereum) dynamically change the complexity of the blocks. However, the RL techniques can be used for optimizing the long-term utility of a blockchain network instead of relying on instantaneous gains. For instance, a multi-armed bandit learning network can be used to identify the complexity of the algorithms based on the scale and other characteristics of the network. Thus, in pursuit of developing a long-term optimal policy, if the average block time is less than the expected block time, then the level of difficulty can be increased. In contrast, if the expected time is less than the average block time, then the level of difficulty can be reduced for mining. Thus, RL techniques can help optimize the performance based on different characteristics of the end-to-end blockchain network [69].

5. Case Study: Minimization of Forking in Blockchain

In this section, we present a case study for optimizing IIoT-blockchain networks using RL techniques. The results provided here can act as fundamental building blocks for future research in RL and blockchain-enabled IIoT networks.

5.1. Problem Formulation

We consider a blockchain-enabled IIoT network consisting of multiple miners and a single communication point, as shown in Figure 5. The communication point is the static infrastructure and associated with miners in the network. The miners are mobile units with computational capabilities for gathering transaction data. The ledger is considered to be located at the communication point, whereby, transaction records are stored as blocks. Whenever a transaction record is stored as a block at the communication point, the block needs to be validated first for confirming the originality of transactions. As per Blockchain 1.0, the communication points can delegate this proof-of-work (mining) computation for validation to the wireless miners. Once each miner completes the proof-of-work, an acknowledgment (ACK) message is sent to the communication point by the miner. The communication point then propagates the ACK message to the other IIoT devices. In principle, the ACK messages received by the communication point must be identical to the order of completion of the task.

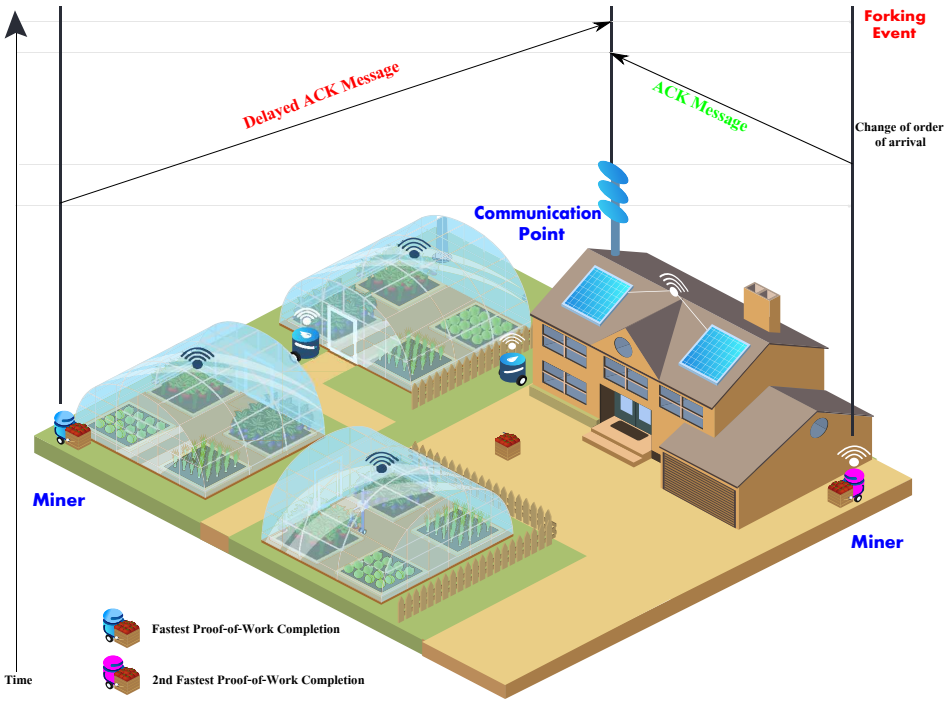


Figure 5. An illustration of the system model. The considered network consists of multiple miners and a single communication point. A forking event takes place when the ACK messages arrive out of order at the communication point.

A forking event can occur in case the ACK first sent by the miner arrives later than other ACKs due to transmission delay [70]. This results in creating branches and the recovery from the forking event increase the overall latency of the network [71]. Note that we consider that K miners and J transaction nodes in the network are equipped with single antennas and communicate with the communication point over the quasi-static flat-fading channel. Additionally, the channel coefficients remain the same during each time slot and vary from a one-time slot to another. The transmission time of the message is the ratio of the size of the message to the overall data rate given by

$$T_t = \frac{\Gamma}{R} \tag{6}$$

where Γ is the size of the message. Here, R is the data rate given as

$$R = \sum_{k=1}^{K} \log\left(1 + \frac{p_k |h_k|^2}{\sum_{j=1}^{J} p_j |h_j| + N_0}\right), \tag{7}$$

where p_k and h_k are the transmit power and channel coefficient of the main link while p_j and h_j are the transmit power and channel coefficient of the interference link. Furthermore, N_0 represents the variance of additive white Gaussian noise (AWGN). Since the size of the message is generally fixed, we aim to maximize the data rate to reduce the transmission delay, thereby reducing the forking event.

5.2. Algorithm Design

To solve the $\max_{p_1,p_2,\ldots,p_K} R$ problem subject to $0 \leq p_k \leq P_m$, we employ a Q-learning technique by controlling the power of the miner. In the following, we define three key elements of the Q-learning model, i.e., states, rewards, and actions.

5.2.1. States

In a decision epoch, a state $s \in \tilde{S}$ is defined by the energy of the miner and the channel gains. The miner takes a decision based on either saving the energy or transferring the message.

5.2.2. Actions

We consider that the miner adaptively switches between energy-saving and message transferring modes in a state. Thus, the action set $a \in \tilde{A}$ can be represented as either 0 or 1, whereby, 0 represents energy-saving and 1 denotes message transferring mode. This is mathematically given as

$$\tilde{A} = \{0, 1\} \tag{8}$$

5.2.3. Reward

Note that the forking event is directly dependent on the successfully transferred information and indirectly depends on the energy of the miner. Thus, the agent receives a reward $\tilde{R}(s^{'}|s, a)$ only when operating in the message transferring mode. In the energy-saving mode, the agent receives no reward but the energy saved can be used for later time slots to support data transmission. Mathematically, the reward function can be expressed as

$$\tilde{R} = R \times a. \tag{9}$$

where a represents the action performed by the agent. It can be noted that when $a = 0$, the agent operates in energy-saving mode and does not receive any reward. During each epoch, the miner chooses an action based on the state-action values and updates the Q-table. The table is initialized by setting Q-value as zero for all the actions and states. The iteration starts with picking a random state and update the Q-table using the $\epsilon-$greedy method. This iterative method aims to maximize the Q-value and the agent receives an immediate reward if the information is transferred successfully. The Q-value is updated as

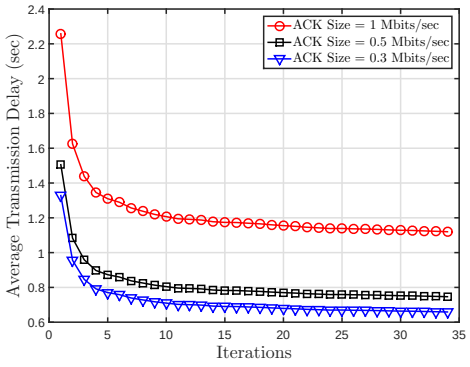
$$Q(s, a) := \alpha \max_{a^{'}} \{\gamma Q(s^{'}, a^{'}) + \tilde{R}(s^{'}|s, a) - Q(s, a)\} + Q(s, a). \tag{10}$$

where α represents the learning rate and γ is the discount factor. The agent makes a better decision over a longer period and the Q-table starts to stabilize. In the end, the mode selection policy is obtained which is based on the best set of actions for different states.

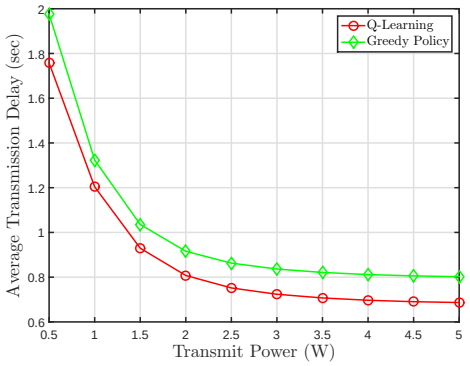
5.3. Results and Discussion

We now evaluate the performance of the proposed Q-learning scheme for reducing the transmission delay of miners and the occurrence of forking events in the blockchain. We performed extensive simulation and created a backscatter communication environment for the agent interaction. The learning agent interacts with the environment and takes an action that maximizes the immediate reward $\tilde{R}$. After performing an action, the agent observes the next state and again interacts with the simulation environment. Over the time and after a specific learning period, the performance starts to converge and the Q-table stabilizes. For the sake of fair comparison, we compare our Q-learning based scheme with a greedy policy. In the case of greedy policy, the miner chooses message transfer mode if it has sufficient energy to transmit the signal. Otherwise, it stays passive with the energy-saving mode.

Figure 6a shows the average transmission delay achieved by using the proposed Q-learning algorithm. Each point in the figure represents the rolling average of the last 10^3 time slots. For each curve, we have illustrated the impact of different ACK message sizes. The smaller message size reduces the transmission delay. Furthermore, the increase in the number of iterations further reduces the transmission delay. This indicates that the agent makes a better decision over time and becomes fairly stable after many interactions with the environment. In Figure 6b, we compare the Q-learning approach with the benchmark greedy policy. It can be seen that the Q-learning approach significantly outperforms the greedy policy. Since the greedy approach only focuses on maximizing the current reward, it performs poorly in terms of average transmission delay. In contrast, the Q-learning approach not only tries to maximize the current reward but also considers future rewards and, thus, optimizes the actions.



(**a**)



(**b**)

Figure 6. Average transmission delay against (**a**) number of iterations (**b**) transmit power.

6. Conclusions and Future Work

Blockchain-enabled IIoT networks are emerging as a new norm in the global development and adoption of blockchain technology. However, the optimization of these networks requires more

than just conventional model-driven methods. Thus, the RL techniques are a viable solution to address the challenges that IIoT-blockchain networks face such as minimizing forking events and improving the transactional throughput. This article has highlighted some of the recent and concrete studies on blockchain-enabled IIoT networks and detailed how RL techniques can be applied to solve the associated issues comprehensively. Subsequently, a case study has been presented where the occurrence of the forking event is minimized using the Q-learning technique. The results demonstrate the improvements offered by the proposed technique over the greedy method.

As a promising candidate for optimizing the performance of blockchain-enabled IIoT networks, the RL techniques are helping to realize intelligence in such networks. However, there are several challenges and open research questions that may drive future research efforts in this domain. Some of the potential research directions include and are not limited to the following:

6.1. Energy Constrained IIoT Devices in Blockchain

One of the major issues in applying the RL techniques to blockchain-enabled IIoT devices is the energy-constrained nature of the IIoT devices. Though it arguably makes the performance of the RL techniques no less significant, it drains the energy of low-powered devices very quickly. Especially, in the case of distributed deep RL techniques, this issue can become more prominent. In these conditions, there is a requirement of energy-efficient management at the physical layer of networks [72].

6.2. Independence of RL Techniques and Blockchain

Typically, the RL and blockchain models operate independently, therefore, the blockchain entities are not able to communicate with the RL processes. This independence can potentially reduce the performance of the blockchain-enabled IIoT network, to the point of jeopardizing the entire infrastructure. Thus, there is a more fervent need for cross-domain research efforts for integrating RL processes within blockchain infrastructure to the level of individual bits and instructions. This embedding is expected to provide improved performance due to higher collective gains for the network.

6.3. The Scalability Paradox

Scalability has been a major area of research in blockchain networks. In recent years, the issue has received much attention for realizing the massive and large-scale blockchain-enabled IIoT networks for smart cities. Although the scalability is an inherent problem of blockchains, this becomes more complex with the introduction of RL techniques. The training for RL techniques needs to be performed extensively and after regular intervals. The deep RL techniques require a more complex architecture of the network for distributed intelligence. The issue worsens when the blockchain-enabled IIoT networks are hyper-mobile, where some devices enter and leave the network recurrently. Thus, there is a need for focused research efforts to make progress in this domain.

6.4. Selection of Appropriate RL Techniques

Another important issue to address is the selection of appropriate RL techniques for optimizing different aspects of the network. Since one size does not fit all, the performance of RL techniques may differ dramatically if applied to the wrong set of problems. As detailed in the above sections, the deep Q-learning method is complex and may be suitable for problems with large state space. On the other hand, less complex techniques like multi-armed bandit learning may be best for addressing small local issues in blockchain-enabled IIoT networks. Due to the high diversity of RL techniques, future research efforts need to focus on developing a compendium of RL techniques for a specific set of problems of blockchain-enabled IIoT networks.

6.5. Financial Forecasting

Some of the recent literature in blockchain domain space has studied and stressed that in many cases, RL and deep RL have better financial forecasting results when compared to supervised learning techniques. For instance, stock price prediction; it is well known that historical data cannot reflect the dynamics of the current market, which contributes to poor prediction performance of future price changes. It is expected that by adopting RL and employing its techniques, better forecasts can be made. Extending this argument to blockchain-enabled IIoT network where hundreds and thousands of agents operate autonomously, it is indispensable for better management that the cost expenses associated with them be predicted in an accurate way.

6.6. Smart Agents

This presents a very interesting research direction, where agents with smart capabilities are designed in a way that they help regulate the blockchain-IIoT network and detect abnormal or malicious behavior patterns with a high probability. Note that the former is especially important for blockchains that are private or consortium in nature, while the latter is required for public blockchains. The design and employment of such agents will not only help regulate blockchains, but it will also help them in achieving self-healing attributes.

6.7. Anonymous Data Sharing

With the recent and ongoing developments in the IIoT space, the issues of privacy are gaining worldwide attention. Blockchain coupled with RL techniques can help enable anonymous data sharing between two users or devices in a blockchain-IIoT network. Thus, multiple-layer structures of blockchain can be designed together with data fusion, which will allow sophisticated authorization of data for different users and enable more complex networks to be designed.

Author Contributions: Conceptualization: F.J., U.J., M.N.A.; Writing Original Draft: F.J., W.U.K., H.P.; Funding Acquisition: R.J.; Supervision: R.J.; Project Administration: R.J. All authors have read and agreed to the published version of the manuscript.

Funding: This work is supported by Business Finland under the project 5G Finnish Open Research Collaboration Ecosystem (5G-FORCE).

Acknowledgments: The work of F. Jameel and R. Jäntti was partly supported by Business Finland under the project 5G Finnish Open Research Collaboration Ecosystem (5G-FORCE).

Conflicts of Interest: The authors declared no conflict of interest.

References

1. Mohanta, B.K.; Jena, D.; Satapathy, U.; Patnaik, S. Survey on IoT Security: Challenges and Solution using Machine Learning, Artificial Intelligence and Blockchain Technology. *Internet Things* **2020**, *11*, 100227. [CrossRef]
2. Javaid, U.; Siang, A.K.; Aman, M.N.; Sikdar, B. Mitigating IoT Device Based DDoS Attacks Using Blockchain. In Proceedings of the 1st Workshop on Cryptocurrencies and Blockchains for Distributed Systems, Munich, Germany, 15 June 2018; pp. 71–76. [CrossRef]
3. Wu, J.; Tran, N.K. Application of Blockchain Technology in Sustainable Energy Systems: An Overview. *Sustainability* **2018**, *10*, 3067. [CrossRef]
4. Sun, Y.; Zhang, L.; Feng, G.; Yang, B.; Cao, B.; Imran, M.A. Blockchain-enabled wireless Internet of Things: Performance analysis and optimal communication node deployment. *IEEE Internet Things J.* **2019**, *6*, 5791–5802. [CrossRef]
5. Alladi, T.; Chamola, V.; Parizi, R.M.; Choo, K.R. Blockchain Applications for Industry 4.0 and Industrial IoT: A Review. *IEEE Access* **2019**, *7*, 176935–176951. [CrossRef]
6. Bansal, K.; Mittal, K.; Ahuja, G.; Singh, A.; Gill, S.S. DeepBus: Machine learning based real time pothole detection system for smart transportation using IoT. *Internet Technol. Lett.* **2020**, *3*, e156. [CrossRef]

7. Jameel, F.; Wyne, S.; Jayakody, D.N.K.; Kaddoum, G.; O'Kennedy, R. Wireless Social Networks: A Survey of Recent Advances, Applications and Challenges. *IEEE Access* **2018**, *6*, 59589–59617. [CrossRef]
8. Jameel, F.; Hamid, Z.; Jabeen, F.; Zeadally, S.; Javed, M.A. A survey of device-to-device communications: Research issues and challenges. *IEEE Commun. Surv. Tutorials* **2018**, *20*, 2133–2168. [CrossRef]
9. Mao, D.; Hao, Z.; Wang, F.; Li, H. Innovative Blockchain-Based Approach for Sustainable and Credible Environment in Food Trade: A Case Study in Shandong Province, China. *Sustainability* **2018**, *10*. [CrossRef]
10. Jameel, F.; Javaid, U.; Sikdar, B.; Khan, I.; Mastorakis, G.; Mavromoustakis, C.X. Optimizing Blockchain Networks with Artificial Intelligence: Towards Efficient and Reliable IoT Applications. In *Convergence of Artificial Intelligence and the Internet of Things*; Springer: Berlin, Germany, 2020; pp. 299–321.
11. Kaur, P.; Singh, A.; Gill, S.S. RGIM: An Integrated Approach to Improve QoS in AODV, DSR and DSDV Routing Protocols for FANETS Using the Chain Mobility Model. *Comput. J.* **2020**. [CrossRef]
12. Jameel, F.; Javed, M.A.; Jayakody, D.N.; Hassan, S.A. On secrecy performance of industrial Internet of things. *Internet Technol. Lett.* **2018**, *1*, e32. [CrossRef]
13. Jameel, F.; Khan, W.U.; Shah, S.T.; Ristaniemi, T. Towards intelligent IoT networks: Reinforcement learning for reliable backscatter communications. In Proceedings of the 2019 IEEE Globecom Workshops (GC Wkshps), Waikoloa, HI, USA, 13 December 2019; pp. 1–6.
14. Alpaydin, E. *Introduction to Machine Learning*; MIT Press: Cambridge, MA, USA, 2009.
15. Jameel, F.; Sharma, N.; Khan, M.A.; Khan, I.; Alam, M.M.; Mastorakis, G.; Mavromoustakis, C.X. Machine Learning Techniques for Wireless-Powered Ambient Backscatter Communications: Enabling Intelligent IoT Networks in 6G Era. In *Convergence of Artificial Intelligence and the Internet of Things*; Springer: Berlin, Germany, 2020; pp. 187–211.
16. massoud Farahmand, A.; Shademan, A.; Jagersand, M.; Szepesvári, C. Model-based and model-free reinforcement learning for visual servoing. In Proceedings of the 2009 IEEE International Conference on Robotics and Automation, Kobe, Japan, 12 May 2009; pp. 2917–2924.
17. Zou, X.; Yang, R.; Yin, C.; Nie, Z.; Wang, H. Deploying tactical communication node vehicles with AlphaZero algorithm. *IET Commun.* **2020**, *14*, 1392–1396. [CrossRef]
18. Ernst, D.; Glavic, M.; Capitanescu, F.; Wehenkel, L. Reinforcement learning versus model predictive control: a comparison on a power system problem. *IEEE Trans. Syst. Man Cybern.* **2008**, *39*, 517–529. [CrossRef] [PubMed]
19. Yang, D.; Roth, H.; Xu, Z.; Milletari, F.; Zhang, L.; Xu, D. Searching Learning Strategy with Reinforcement Learning for 3D Medical Image Segmentation. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Shenzhen, China, 17 October 2019; pp. 3–11.
20. Arulkumaran, K.; Deisenroth, M.P.; Brundage, M.; Bharath, A.A. Deep reinforcement learning: A brief survey. *IEEE Signal Process. Mag.* **2017**, *34*, 26–38. [CrossRef]
21. Brewer, E.A. Towards Robust Distributed Systems (Abstract). In Proceedings of the Nineteenth Annual ACM Symposium on Principles of Distributed Computing, Portland, OR, USA, 16 July 2000. [CrossRef]
22. Sun, Q.; Li, H.; Ma, Z.; Wang, C.; Campillo, J.; Zhang, Q.; Wallin, F.; Guo, J. A Comprehensive Review of Smart Energy Meters in Intelligent Energy Networks. *IEEE IoT J.* **2016**, *3*, 464–479. [CrossRef]
23. Lei, A.; Cao, Y.; Bao, S.; Li, D.; Asuquo, P.; Cruickshank, H.; Sun, Z. A blockchain based certificate revocation scheme for vehicular communication systems. *Future Gener. Comput. Syst.* **2019**, *110*, 892–903. [CrossRef]
24. Ren, Y.; Leng, Y.; Zhu, F.; Wang, J.; Kim, H.J. Data Storage Mechanism Based on Blockchain with Privacy Protection in Wireless Body Area Network. *Sensors* **2019**, *19*, 2395. [CrossRef]
25. Anderson, R.; Fuloria, S. Who Controls the off Switch? In Proceedings of the 2010 First IEEE Int. Conference on Smart Grid Communication, Gaithersburg, MD, USA, 2010; pp. 96–101. [CrossRef]
26. Wang, X.; Zha, X.; Ni, W.; Liu, R.P.; Guo, Y.J.; Niu, X.; Zheng, K. Survey on blockchain for Internet of Things. *Comput. Commun.* **2019**, *136*, 10–29. [CrossRef]
27. Duy, P.T.; Hien, D.T.T.; Hien, D.H.; Pham, V.H. A survey on opportunities and challenges of Blockchain technology adoption for revolutionary innovation. In Proceedings of the Ninth International Symposium on Information and Communication Technology, Lahore, Pakistan, 12 November 2018; pp. 200–207.
28. Liu, J.; Liu, Z. A survey on security verification of blockchain smart contracts. *IEEE Access* **2019**, *7*, 77894–77904. [CrossRef]
29. Joshi, A.P.; Han, M.; Wang, Y. A survey on security and privacy issues of blockchain technology. *Math. Found. Comput.* **2018**, *1*, 121–147. [CrossRef]

30. Yang, R.; Yu, F.R.; Si, P.; Yang, Z.; Zhang, Y. Integrated blockchain and edge computing systems: A survey, some research issues and challenges. *IEEE Commun. Surv. Tutorials* **2019**, *21*, 1508–1532. [CrossRef]
31. Xie, J.; Yu, F.R.; Huang, T.; Xie, R.; Liu, J.; Liu, Y. A Survey on the Scalability of Blockchain Systems. *IEEE Network* **2019**, *33*, 166–173. [CrossRef]
32. Atzei, N.; Bartoletti, M.; Cimoli, T. A survey of attacks on ethereum smart contracts (sok). In Proceedings of the International conference on principles of security and trust, Prague, Czech Republic, 11 April 2017; pp. 164–186.
33. Rouhani, S.; Deters, R. Security, performance, and applications of smart contracts: A systematic survey. *IEEE Access* **2019**, *7*, 50759–50779. [CrossRef]
34. Li, X.; Jiang, P.; Chen, T.; Luo, X.; Wen, Q. A survey on the security of blockchain systems. *Future Gener. Comput. Syst.* **2020**, *107*, 841–853. [CrossRef]
35. Danzi, P.; Kalor, A.E.; Stefanovic, C.; Popovski, P. Analysis of the communication traffic for blockchain synchronization of IoT devices. In Proceedings of the 2018 IEEE International Conference on Communications (ICC), Kansas City, MO, USA, 24 May 2018; pp. 1–7.
36. Jabeen, T.; Ali, Z.; Khan, W.U.; Jameel, F.; Khan, I.; Sidhu, G.A.S.; Choi, B.J. Joint Power Allocation and Link Selection for Multi-Carrier Buffer Aided Relay Network. *Electronics* **2019**, *8*, 686. [CrossRef]
37. Sun, Y.; Zhang, L.; Feng, G.; Yang, B.; Cao, B.; Imran, M. Performance Analysis for Blockchain Driven Wireless IoT Systems Based on Tempo-Spatial Model. In Proceedings of the 2019 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC), Guilin, China, 15 June 2019; pp. 348–353.
38. Aitzhan, N.Z.; Svetinovic, D. Security and Privacy in Decentralized Energy Trading Through Multi-Signatures, Blockchain and Anonymous Messaging Streams. *IEEE Trans. Dependable Secur. Comput.* **2018**, *15*, 840–852. [CrossRef]
39. Jameel, F.; Chang, Z.; Huang, J.; Ristaniemi, T. Internet of Autonomous Vehicles: Architecture, Features, and Socio-Technological Challenges. *IEEE Wirel. Commun.* **2019**, *26*, 21–29. [CrossRef]
40. Jameel, F.; Hamid, Z.; Jabeen, F.; Javed, M.A. Impact of co-channel interference on the performance of VANETs under α-μ fading. *AEU Int. J. Electron. Commun.* **2018**, *83*, 263–269. [CrossRef]
41. Jameel, F.; Duan, R.; Chang, Z.; Liljemark, A.; Ristaniemi, T.; Jantti, R. Applications of Backscatter Communications for Healthcare Networks. *IEEE Netw.* **2019**, *33*, 50–57. [CrossRef]
42. Rifi, N.; Agoulmine, N.; Chendeb Taher, N.; Rachkidi, E. Blockchain technology: Is it a good candidate for securing iot sensitive medical data? *Wirel. Commun. Mob. Comput.* **2018**, *2018*, 9763937. [CrossRef]
43. Zhang, R.; Xue, R.; Liu, L. Security and Privacy on Blockchain. *ACM Comput. Surv.* **2019**, *52*, 51:1–51:34. [CrossRef]
44. Burhan, M.; Rehman, R.; Khan, B.; Kim, B.S. IoT elements, layered architectures and security issues: A comprehensive survey. *Sensors* **2018**, *18*, 2796. [CrossRef] [PubMed]
45. Javaid, U.; Aman, M.N.; Sikdar, B. BlockPro: Blockchain Based Data Provenance and Integrity for Secure IoT Environments. In Proceedings of the 1st Workshop on Blockchain-enabled Networked Sensor Systems, Shenzhen, China, 10 November 2018; pp. 13–18. [CrossRef]
46. Javaid, U.; Aman, M.N.; Sikdar, B. DrivMan: Driving Trust Management and Data Sharing in VANETs with Blockchain and Smart Contracts. In Proceedings of the 2019 IEEE 89th Vehicular Technology Conference (VTC2019-Spring), Kuala Lumpur, Malaysia, 1 May 2019; pp. 1–5.
47. Jiang, L.; Xie, S.; Maharjan, S.; Zhang, Y. Blockchain Empowered Wireless Power Transfer for Green and Secure Internet of Things. *IEEE Netw.* **2019**, *33*, 164–171. [CrossRef]
48. Watanabe, H.; Fan, H. A Novel Chip-Level Blockchain Security Solution for the Internet of Things Networks. *Technologies* **2019**, *7*. [CrossRef]
49. Chaintoutis, C.; Akriotou, M.; Mesaritakis, C.; Komnios, I.; Karamitros, D.; Fragkos, A.; Syvridis, D. Optical PUFs as physical root of trust for blockchain-driven applications. *IET Softw.* **2019**, *13*, 182–186. [CrossRef]
50. Gill, S.S.; Tuli, S.; Xu, M.; Singh, I.; Singh, K.V.; Lindsay, D.; Tuli, S.; Smirnova, D.; Singh, M.; Jain, U.; et al. Transformative effects of IoT, Blockchain and Artificial Intelligence on cloud computing: Evolution, vision, trends and open challenges. *Internet Things* **2019**, *8*, 100118. [CrossRef]
51. Rawat, D.B. Fusion of Software Defined Networking, Edge Computing, and Blockchain Technology for Wireless Network Virtualization. *IEEE Commun. Mag.* **2019**, *57*, 50–55. [CrossRef]
52. Ioini, N.E.; Pahl, C. A Review of Distributed Ledger Technologies. On the Move to Meaningful Internet Systems. In Proceedings of the OTM 2018 Conferences—Confederated International Conferences: CoopIS, C&TC, and ODBASE 2018, Valletta, Malta, 22–26 October 2018; pp. 277–288. [CrossRef]

53. Hu, Y.; Manzoor, A.; Ekparinya, P.; Liyanage, M.; Thilakarathna, K.; Jourjon, G.; Seneviratne, A. A Delay-Tolerant Payment Scheme Based on the Ethereum Blockchain. *IEEE Access* **2019**, *7*, 33159–33172. [CrossRef]
54. Li, A.; Wei, X.; He, Z. Robust Proof of Stake: A New Consensus Protocol for Sustainable Blockchain Systems. *Sustainability* **2020**, *12*, 2824. [CrossRef]
55. Borkowski, M.; Sigwart, M.; Frauenthaler, P.; Hukkinen, T.; Schulte, S. Dextt: Deterministic Cross-Blockchain Token Transfers. *IEEE Access* **2019**, *7*, 111030–111042. [CrossRef]
56. Kumar, N.M.; Mallick, P.K. Blockchain technology for security issues and challenges in IoT. *Procedia Comput. Sci.* **2018**, *132*, 1815–1823. [CrossRef]
57. Li, J.; Ni, X.; Yuan, Y. The Reserve Price of Ad Impressions in Multi-Channel Real-Time Bidding Markets. *IEEE Trans. Comput. Soc. Syst.* **2018**, *5*, 583–592. [CrossRef]
58. Qian, Y.; Jiang, Y.; Chen, J.; Zhang, Y.; Song, J.; Zhou, M.; Pustišek, M. Towards decentralized IoT security enhancement: A blockchain approach. *Comput. Electr. Eng.* **2018**, *72*, 266–273. [CrossRef]
59. Griggs, K.N.; Ossipova, O.; Kohlios, C.P.; Baccarini, A.N.; Howson, E.A.; Hayajneh, T. Healthcare blockchain system using smart contracts for secure automated remote patient monitoring. *J. Med. Syst.* **2018**, *42*, 130. [CrossRef] [PubMed]
60. Khan, M.A.; Salah, K. IoT security: Review, blockchain solutions, and open challenges. *Future Gener. Comput. Syst.* **2018**, *82*, 395–411. [CrossRef]
61. Park, L.W.; Lee, S.; Chang, H. A sustainable home energy prosumer-chain methodology with energy tags over the blockchain. *Sustainability* **2018**, *10*, 658. [CrossRef]
62. Liu, M.; Teng, Y.; Yu, F.R.; Leung, V.C.; Song, M. Deep Reinforcement Learning Based Performance Optimization in Blockchain-Enabled Internet of Vehicle. In Proceedings of the ICC 2019-2019 IEEE International Conference on Communications (ICC), Shanghai, China, 24 May 2019; pp. 1–6.
63. Herrera-Joancomartí, J.; Pérez-Solà, C. Privacy in bitcoin transactions: new challenges from blockchain scalability solutions. In Proceedings of the International Conference on Modeling Decisions for Artificial Intelligence, Sant Julia de Loria, Andorra, 19 September 2016; pp. 26–44.
64. Thakkar, P.; Nathan, S.; Viswanathan, B. Performance benchmarking and optimizing hyperledger fabric blockchain platform. In Proceedings of the 2018 IEEE 26th International Symposium on Modeling, Analysis, and Simulation of Computer and Telecommunication Systems (MASCOTS), Milwaukee, WI, USA, 28 Septemebr 2018; pp. 264–276.
65. Yli-Huumo, J.; Ko, D.; Choi, S.; Park, S.; Smolander, K. Where is current research on blockchain technology? A systematiC review. *PLoS ONE* **2016**, *11*, e0163477. [CrossRef] [PubMed]
66. Luong, N.C.; Anh, T.T.; Binh, H.T.T.; Niyato, D.; Kim, D.I.; Liang, Y.C. Joint transaction transmission and channel selection in cognitive radio based blockchain networks: A deep reinforcement learning approach. In Proceedings of the ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brighton, UK, 17 May 2019; pp. 8409–8413.
67. Xu, H.; Zhang, L.; Liu, Y.; Cao, B. Raft based wireless blockchain networks in the presence of malicious jamming. *IEEE Wirel. Commun. Lett.* **2020**, *9*, 817–821. [CrossRef]
68. Jameel, F.; Wyne, S.; Kaddoum, G.; Duong, T.Q. A comprehensive survey on cooperative relaying and jamming strategies for physical layer security. *IEEE Commun. Surv. Tutorials* **2018**, *21*, 2734–2771. [CrossRef]
69. Liu, M.; Yu, F.R.; Teng, Y.; Leung, V.C.; Song, M. Performance optimization for blockchain-enabled industrial internet of things (IIoT) systems: A deep reinforcement learning approach. *IEEE Trans. Ind. Informatics* **2019**, *15*, 3559–3570. [CrossRef]
70. Danzi, P.; Kalør, A.E.; Stefanović, Č.; Popovski, P. Delay and communication tradeoffs for blockchain systems with lightweight IoT clients. *IEEE Internet Things J.* **2019**, *6*, 2354–2365. [CrossRef]
71. Alrubei, S.; Ball, E.; Rigelsford, J.; Willis, C. Latency and Performance Analyses of Real-World Wireless IoT-Blockchain Application. *IEEE Sensors J.* **2020**, *13*, 7372–7383. [CrossRef]
72. Novo, O. Blockchain meets IoT: An architecture for scalable access management in IoT. *IEEE Internet Things J.* **2018**, *5*, 1184–1195. [CrossRef]