LS4BUCC: A Low Overhead Storage Architecture for Blockchain Based Unmanned Collaborative Cognition System

Xunhui Zhang, Huaimin Wang, Peichang Shi and Xiang Fu
College of Computer Science and Technology
National University of Defense Technology
Changsha, Hunan, China
{zhangxunhui, hmwang}@nudt.edu.cn, shipeichang@kylinos.cn, fuxiang13@nudt.edu.cn

***Abstract*—In unmanned collaborative cognition scenarios, traditional IoT systems are facing with data security and node privacy problems. However, blockchain together with data encryption technique enables the trustless network due to the tamper-proof, anonymity and unbreakable properties. Nevertheless, there are still some shortages when combining blockchain with IoT systems, for example the storage consumption. Generally speaking, validation nodes in blockchain network need to store a complete copy of the ledger so as to verify transactions and reach the whole network consensus. However, in unmanned collaborative cognition scenarios, many devices' storage capacity are limited. Therefore, in order to solve the storage problem, we propose *LS4BUCC*, a low overhead storage architecture regarding to the new features of unmanned collaborative cognition systems. It selectively stores data according to the value density. Based on three storage mechanisms and two supportive methods, *LS4BUCC* can reduce the data storage and maintain the system security and efficiency at the same time.**

Keywords-Unmanned Collaborative Cognition; IoT System; Peer-to-peer Network; Blockchain; Low Overhead Storage

I. Motivation and Background

A. Unmanned Collaborative Cognition System

With the rapid development of Internet and the increasement of intelligence for smart devices, unmanned collaborative cognition (UCC) system gradually becomes prosperous, for example the unmanned aerial vehicle (UAV) monitoring system. UCC system is actually an IoT system in special scenarios, where peers interconnect with each other. Through individual perception and cooperation, they share their local cognitions, and finally form into a unified cognition. Thereafter, the system makes decisions according to the current cognition and historical data.

The features of UCC system:

- *Uneven storage capacity of nodes*. Just like traditional IoT system, the device's performance and storage capacity are uneven. Many devices are limited by their power dissipation, production cost or physical space. However, with the development of flash memory techniques, there are still many strong nodes in the whole network.
- *Low correlation of current and historical perceptual data*. The actions of peers in UCC system are divided by tasks or missions. There is new perceptual data under new circumstances.
- *Frequently changing situations*. One of the key parts of UCC system is environmental perception. However, the perceiver, the perceived object and the environment will change dynamically, which leads to the fluctuation of the whole situation.
- *Complex and diverse perceptual data*. There are many types of perceptual data, including geographical location information, object state information, weather information and etc. And the acquired information changes frequently due to the change of situations.

B. The Combination of Blockchain and IoT

IoT systems, including UCC system are still facing big challenges, including privacy protection, cyber attacks, physical tampering and etc. The generation of blockchain makes it possible to solve the above mentioned problems due to its tamper-proof, anonymity and unbreakable properties [1]. And the combination of blockchain and IoT system has been widly used in many areas, including electronic business [2], smart home [3] and etc.

However, blockchain brings about new challenges for IoT system, one of which is the storage of ledger due to the uneven storage capacity of system nodes. By the end of September 2018, the size of bitcoin blockchain has reached approximately 184 gigabytes [4], and is still increasing rapidly. Meanwhile, the Moore's law has also hit the bottleneck [5], and the speed of hardware development slows down. The rapidly increasing amount of data and the low speed increasing data capacity together hinder the application of blockchain in IoT system. Moreover, for UCC, the frequently changing situation also increases the storage burden. Therefore, in this paper, we will mainly consider the solution of data storage in blockchain based UCC systems.

C. New Features of Blockchain Based UCC

Apart from the storage problem, there are some new features of UCC that we need to take into consideration when combining with blockchain.

978-1-7281-1442-2/19/$31.00 ©2019 IEEE
DOI 10.1109/SOSE.2019.00038

- The data transmission among nodes and the reoccurrence of local complete semantics. Due to the limitation of network bandwith, we need to simplify the transferred data so that we can ensure the efficiency. However, in order to understand the information obtained from other nodes, we need to design an approach in advance to ensure that a node can generate the complete information from simplified data.
- The value density of historical information. Even though there is low correlation of current and historical perceptual data in UCC systems, especially for different tasks' data. Still the past empirical data may guide to make decisions and plans. Therefore, the system needs to store data according to its value density. Meanwhile, the system needs to measure the value density automatically.
- The particularity of consensus information. Unlike Bitcoin [1], Ethereum [2] and other cryptocurrencies, the consensus information of UCC system is not the trade information, which needs to be verified via the context information. Instead, nodes only need to verify the correctness of the current information and the validity of sender.

By taking the new features mentioned above into consideration, we come up with a solution for the storage of blockchain based UCC system.

D. Improvement Strategy

Firstly, the low correlation property makes it feasible to partition storage data according to tasks, whereas we need to ensure that the historical data is still accessible so as to support rediscussion. Secondly, there are differences between the value densities of historical data, and it is suitable to store them differently while considering the storage capacity at the same time. Thirdly, in UCC system, nodes can generate rich semantic information due to the complex and diverse perceptual data. However, in order to reduce the data storage and data transfer, nodes can use the preset semantic template to simplify information.

Due to the problems and strategies mentioned above, we consider designing a low overhead data storage architecture for blockchain based UCC system. Meanwhile, we propose three mechanisms and two supportive methods to adapt to the storage architecture and the new features of UCC system mentioned in section I-C.

The contributions of this paper are as follows:

- We propose *LS4BUCC* model for low overhead storage of blockchain based UCC system.
- We come up with three mechanisms to reduce the storage overhead of the proposed architecture based on the data relevance.
- We propose two preliminary methods to support the three mechanisms. One is to estimate the value density of data, the other one is to selectively store historical data more effectively.

The rest of this paper are organized as follows. Section 2 reviews some related studies. Section 3 proposes the architecture of *LS4BUCC*. Section 4 discusses the effectiveness of our model. Finally, in section 5, we present the conclusion and future work.

II. Related Work

In this part, we will firstly present some related works for IoT to see its development and challenges. Thereafter, we show the development of blockchain. Finally, we will present the related work about the combination of blockchain and IoT.

A. Development of IoT

As early as 2008, US National Intelligence Council had regarded IoT as one of the "Disruptive Civil Technology", and considered that by 2025, Internet nodes would reside in everyday things [6]. Nowadays, IoT has been applied in many domains, including personal health care, smart home, agriculture and etc. Luo et al. [7] proposed a cost-effective health case system through wireless network. Alkar et al. [8], Darianian et al. [9], and Kelly et al. [10] used IoT concept in smart home to control the equipments and monitor environment conditions. Zhao et al. [11] proposed an agricultural application of wirless sensor network, through which provides scientific guidance for agricultural production. However, due to the storage and computation limitations, IoT always resort to cloud services [12]. Therefore, IoT's features leads to the security and privacy problems, including lack of central control, heterogeneity in device resources, scalibility and etc. [13].

B. Development of Blockchain

Since 2008, Nakamoto proposed bitcoin in the whitepaper [14], blockchain came into being. Its decentralization, persistency, anonymity and auditability made it be widly used from then on [15]. Bitcoin, Ethereum, Hyperledger and etc. were proposed as cryptocurrencies, which brought about a profound impact on traditional finance. Devine [16] proposed blockchain learning, and Akins et al. [17] came up with an income tax system, which showed the blockchain's impact on social services. Moreover, blockchain can be used to enhance security and protect privacy. Aitzhan et al. [18] solved the security and privacy problem in decentralized energy trading system by using blockchain. Dorri et al. [19] used blockchain to protect the security and privacy in smart vehicles.

C. Convergence of Blockchain and IoT

According to the above related works, blockchain may solve the security and privacy problem that traditional IoT faces to some extent. Just as Kshetri said [5], the conbination of blockchain and IoT would improve the security of IoT to some extent. It could not only solve the privacy management problem in IoT, but also defense against cyber attacks and software bugs compared to centralized cloud service. Samaniego et al. [20] also considered that blockchain could serve IoT, and

[1] https://bitcoin.org/en/

[2] https://www.ethereum.org/

verified the performance of the conbination. The result showed that the conbination of blockchain and IoT ourperformed cloud service. Because of blockchain's characteristics mentioned above, it can bring about the turnaround for IoT, and it has already been used in many IoT applications to improve service quality. Liu et al. [21] proposed a blockchain based data integrity verification service, which could tolerate the dynamic property of IoT by depending on the third party auditor. Dori et al. [3] designed a smart home architecture based on blockchain, which implemented action control and access management by modifying block's header information. Zhang et al. [2] and Wörner et al. [22] proposed effective use of blockchain in the electric business and data exchange respectively. For the development, testing and fault tolerance analysis of IoT blockchain applications, Walker et al. [23] designed PlaTIBART.

However, it is undeniable that there are still many challenges in the convergence of blockchain and IoT. Just as Dorri et al. [13] said, even though blockchain can handle most security and privacy threats of IoT, the decentralization and resource-constraints are still challenges. Zyskind et al. [24] also considered that blockchain based distributed storage would lead to huge redundance. They proposed Enigma, which used off-chain storage. By using DHT protocol, it removed duplications across network regarding to data content. Nevertheless, it can only reduce the size of off-chain data. There is still a large scale and fast-growing on-chain data.

In this paper, we mainly focus on the solution of storage problem regarding to UCC system. Next we will introduce our proposed architecture in detail.

III. Architecture

In this section, we will describe the architecture of *LS4BUCC* as shown in figure 1. Aiming at solving the problem of UCC system's data surpasses node's storage capacity. By taking data redundancy, the change of data value according to the occurrence time, and the relevance of historical data into consideration comprehensively, we propose the low overhead storage model for blockchain based unmanned collaborative cognition system.

There are mainly three mechanisms, namely the *Semantic Information Template* mechanism, the *Hysteretic Data Slice* mechanism, and the *Archive Historical Data* mechanism. The node will judge which mechanism to use according to the value density of the data. Meanwhile, there are two methods that support the three mechanisms, namely *Value Density Method* and *Selective Storage Method*.

Next, we will firstly introduce the two methods that support the three mechanisms.

A. Supportive Methods

1) Value Density Method: The value density represents the importance of a piece of data for the target node in UCC system. Mostly, the perceptual and decision-making data are used to support making future decisions or plans.

$$VD(data, node) \propto \frac{logN}{logT} \tag{1}$$

As shown in equation 1, where VD means value density, N represents the number of times that the data has been used by the target node, and T represents the time since the data generated. The value density is proportional to the logarithmic value of N, and is inversely proportional to the logarithmic value of T. To judge whether a node will store a piece of data, it needs to consider not only the value density, but also its own storage capacity.

2) Selective Storage Method: In order to reduce the storage burden of nodes, one need to selectively discard and store the data that is already existed in the storage space. However, it is a problem to decide what data to store. If all the nodes select to discard a piece of data, this will leads to data loss in the beginning. If randomly decide whether to store or not, this will lead to uneven data distribution. Some data will be stored massively, however some other data will be stored in few or even no node. Therefore, we come up with a selective storage method according to the Kademlia technique [25]. Kademlia is a popular peer-to-peer distributed hash table (DHT), which performs better than Chord, CAN, Pastry and other DHT techniques to some extent. BitTorrent [3] also implemented a file transfer protocol based on Kademlia.

Kademlia uses the XOR metric to calculate the logic distance between each pair of nodes or between a node and the hash value of data. Meanwhile, the data information can also get the corresponding hash value by hash algorithm. Therefore, we can design a method to store a piece of data in its k relatively close nodes according to the logic distance because we also need to take nodes' storage capacity into consideration. And we use Kademlia to maintain the DHT in each node. In this way, we can reduce the data storage under the premise of data balance.

B. Mechanisms for LS4BUCC

Semantic Information Template. The sensors of equipment firstly get the environmental data through perception, and form into a piece of simplified semantic infomation through build-in methods. Also, it can get the simplified data from other nodes. The self-acquired data will also be transferred to other nodes to reach consensus through specific consensus protocol. After storing the simplified semantic information into local cache memory, the node needs to generate the whole semantic information so as to support the make of decision. The decision and plan should also reach consensus among related nodes. For the introduction of semantic information template, we take the UAV monitoring system as an example. As shown in table I, the simplified semantic information is generated based on the gathered perceptual data and simplified data from other nodes. And the corresponding whole semantic information can be seen in figure 2. From this pair of simplified and detailed information, we can see that the templates of semantic information are formed by agreed-on rules. And the templates can not only reduce the data storage and transfer, but also encrypt the data and improve the system security.

[3] https://www.bittorrent.com/

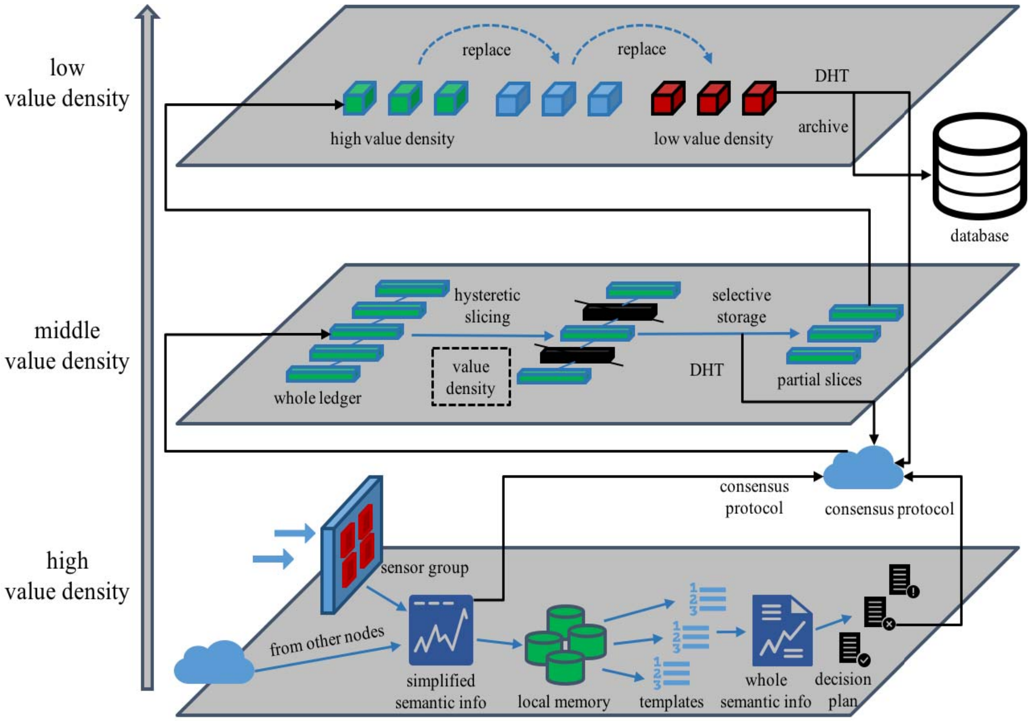

Fig. 1: Architecture for *LS4BUCC*

TABLE I: Example for UAV Original Data

O (4541, 1806, 1488)
L (113.010435, 28.217831, 40)
A (M, 13.9, 1, N)
E (30.2, 45)
. . .

There is an **O**bject, which is **4541**mm in length, **1806**mm in width, and **1488**mm in height.

The object's **L**ocation is: longitude -- **113.010435**, latitude -- **28.217831**, altitude -- **40**m.

The **A**ction of the object is **M**oving to the **N**orth with speed **13.9**m/s and acceleration **1**m/s^2.

For the surrounding **E**nvironment, its ambient temperature is **30.2**°F, and its ambient humidity is **45**%RH.

...

Fig. 2: Example for Detailed Information of UAV

Hysteretic Data Slice. In order to reduce the storage pressure of each node, it firstly slice the ledger into slices. Then store different slices into different nodes according to the data value density and node storage capacity dynamically. If a slice of data has high value density, it may be stored locally. Otherwise, it will be stored in its nearest k nodes based on the selective storage method. In this way, the nodes can reduce the storage costs. Moreover, the distribution of ledger can increase the security of the whole system, which makes the invader harder to find the whole ledger.

Archive Historical Data. In UCC systems, if the data's value density is very low, and the related node's storage capacity does not allow to store more data, then the low value density data will be archived in centralized database.

These are the mechanisms used in *LS4BUCC*, which brings about the reduce of storage in UCC systems. However, it is inevitable that some nodes may need to get extra data that is not stored locally. Therefore, the data retrieval and acquisition is also important in the whole peer-to-peer system. In short, there are three steps when getting data.

1) Check the local store list, and get the related high value density data.
2) Recursively search the DHT in related peers, and get the data in logically near node according to Kademlia technique.
3) Search the central database for the target data.

Most often, step 2 and step 3 start simultaneously, because it is unknown which way is faster. However, to balance the workload of center node and to ensure the whole system can function well when facing single point of failure, it is necessary to design such storage and query mechanism.

IV. Discussion

A. Reduction of Storage

Generally speaking, in order to verify transactions and reach consensus, validation nodes need to store the whole ledger, such as Bitcoin, Ethereum. And many personal users cannot become full nodes because of their storage capacity, for example, using Bitcoin wallet on cellphone. However, in *LS4BUCC*, we use three mechanisms to reduce the storage information. For *Semantic Information Template* mechanism, it simplifies the original data to several keywords. For *Hysteretic Data Slice* mechanism, it divides the whole ledger into slices. Each slice is only stored in k ($k \leq node\,number$) nodes. For *Archive Historical Data* mechanism, all the low value density data are removed from peer nodes and stored into central database. In conclusion, based on the three mechanisms, *LS4BUCC* can reduce the data storage on nodes.

B. System Security

The methods and mechanisms mentioned above can improve the system security and robustness to some extent. Firstly, the *Semantic Information Template* is actually a kind of data encryption. The simplified data cannot be recovered easily without the predefined template. Secondly, the Kademlia DHT technique uses randomly generated 160-bit identifier, which makes the nodes distributed geographically. Therefore, the data can hardly lose even facing attack. Moreover, the k backups of data also guarantee the security of data comparing to random discard. Therefore, *LS4BUCC* can ensure the system security to some extent regarding to the UCC scenario.

C. System Efficiency

Even though *LS4BUCC* is not same as other traditional blockchain, which save the whole ledger in validation nodes, it can still reach a relatively high efficiency when retrieving related information. For Kademlia DHT technique, the time complexity of data lookup is $O(logn)$, where n is the number of nodes in the whole network. Meanwhile, the query from central database can cooperate with DHT at the same time under normal conditions. Moreover, nodes will store part of historical data redundantly according to the value density, which accelerate the decision making process. Therefore, the system efficiency of *LS4BUCC* can be guaranteed to a large extent.

V. Conclusion and Future Work

In UCC scenarios, blockchain has become an effective technique to improve system security and node privacy. However, the traditional way of data storage put forward high requirement to nodes in the system, which is not suitable for most UCC scenarios. Therefore, in this paper, we proposed *LS4BUCC*, a low overhead storage architecture for blockchain based unmanned collaborative cognition system, in which data can be treated differently according to its value density to the related node. There are three mechanisms in the architecture, namely *Semantic Information Template* mechanism, *Hysteretic Data Slice* mechanism, and *Archive Historical Data* mechanism. And we proposed two preliminary supportive methods, *Value Density Method*, and *Selective Storage Method*. Through these mechanisms and methods, we reduce the data storage of nodes under the premise of system security and efficiency, which improves the scalability of UCC system.

However, as we have mentioned before, the consensus information of UCC system is different to traditional trade transactions. Also, the consensus information in this architecture involves perceptual data and decision information. Moreover, the frequently changing situations in UCC scenario lead to high requirements for the throughput of consensus protocol. Therefore, we need to design novel consensus algorithms for this architecture. In the future, we will continuously improve the performance of *LS4BUCC*, designing a suitable consensus protocol and improving the DHT method for higher query efficiency.

Acknowledgment

The research is supported by the National Key R&D Program of China (Grant No.2016YFB1000100), National Natural Science Foundation of China (Grant No.61772030), and the GF Innovative Research Program.

References

[1] A. Banafa, “Iot and blockchain convergence: Benefits and challenges,” *IEEE Internet of Things*, 2017.

[2] Y. Zhang and J. Wen, “The iot electric business model: Using blockchain technology for the internet of things,” *Peer-to-Peer Networking and Applications*, vol. 10, no. 4, pp. 983–994, 2017.

[3] A. Dorri, S. S. Kanhere, R. Jurdak, and P. Gauravaram, “Blockchain for iot security and privacy: The case study of a smart home,” in *Pervasive Computing and Communications Workshops (PerCom Workshops), 2017 IEEE International Conference on*. IEEE, 2017, pp. 618–623.

[4] A. T. Sherman, F. Javani, H. Zhang, and E. Golaszewski, “On the origins and variations of blockchain technologies,” *arXiv preprint arXiv:1810.06130*, 2018.

[5] L. B. Kish, “End of moore’s law: thermal (noise) death of integration in micro and nano electronics,” *Physics Letters A*, vol. 305, no. 3-4, pp. 144–149, 2002.

[6] S. Intelligence, “Six technologies with potential impacts on us interests out to 2025,” *National Intelligent Concil, Tech. Rep*, 2008.

[7] H. Luo, S. Ci, D. Wu, N. Stergiou, and K.-C. Siu, “A remote markerless human gait tracking for e-healthcare based on content-aware wireless multimedia communications,” *IEEE Wireless Communications*, vol. 17, no. 1, 2010.

[8] A. Z. Alkar and U. Buhur, “An internet based wireless home automation system for multifunctional devices,” *IEEE Transactions on Consumer Electronics*, vol. 51, no. 4, pp. 1169–1174, 2005.

[9] M. Darianian and M. P. Michael, “Smart home mobile rfid-based internet-of-things systems and services,” in *Advanced Computer Theory and Engineering, 2008. ICACTE’08. International Conference on*. IEEE, 2008, pp. 116–120.

[10] S. D. T. Kelly, N. K. Suryadevara, and S. C. Mukhopadhyay, “Towards the implementation of iot for environmental condition monitoring in homes,” *IEEE Sensors Journal*, vol. 13, no. 10, pp. 3846–3853, 2013.

[11] Z. Liqiang, Y. Shouyi, L. Leibo, Z. Zhen, and W. Shaojun, “A crop monitoring system based on wireless sensor network,” *Procedia Environmental Sciences*, vol. 11, pp. 558–565, 2011.

[12] J. Zhou, Z. Cao, X. Dong, and A. V. Vasilakos, “Security and privacy for cloud-based iot: challenges,” *IEEE Communications Magazine*, vol. 55, no. 1, pp. 26–33, 2017.

[13] A. Dorri, S. S. Kanhere, and R. Jurdak, “Blockchain in internet of things: challenges and solutions,” *arXiv preprint arXiv:1608.05187*, 2016.

[14] S. Nakamoto, “Bitcoin: A peer-to-peer electronic cash system,” 2008.

[15] Z. Zheng, S. Xie, H.-N. Dai, and H. Wang, "Blockchain challenges and opportunities: A survey," *Work Pap.–2016*, 2016.

[16] P. Devine, "Blockchain learning: can crypto-currency methods be appropriated to enhance online learning?" 2015.

[17] B. W. Akins, J. L. Chapman, and J. M. Gordon, "A whole new world: Income tax considerations of the bitcoin economy," *Pitt. Tax Rev.*, vol. 12, p. 25, 2014.

[18] N. Z. Aitzhan and D. Svetinovic, "Security and privacy in decentralized energy trading through multi-signatures, blockchain and anonymous messaging streams," *IEEE Transactions on Dependable and Secure Computing*, vol. 15, no. 5, pp. 840–852, 2018.

[19] A. Dorri, M. Steger, S. S. Kanhere, and R. Jurdak, "Blockchain: A distributed solution to automotive security and privacy," *IEEE Communications Magazine*, vol. 55, no. 12, pp. 119–125, 2017.

[20] M. Samaniego and R. Deters, "Blockchain as a service for iot," in *Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData), 2016 IEEE International Conference on*. IEEE, 2016, pp. 433–436.

[21] B. Liu, X. L. Yu, S. Chen, X. Xu, and L. Zhu, "Blockchain based data integrity service framework for iot data," in *Web Services (ICWS), 2017 IEEE International Conference on*. IEEE, 2017, pp. 468–475.

[22] D. Wörner and T. von Bomhard, "When your sensor earns money: exchanging data for cash with bitcoin," in *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication*. ACM, 2014, pp. 295–298.

[23] M. A. Walker, A. Dubey, A. Laszka, and D. C. Schmidt, "Platibart: a platform for transactive iot blockchain applications with repeatable testing," in *Proceedings of the 4th Workshop on Middleware and Applications for the Internet of Things*. ACM, 2017, pp. 17–22.

[24] G. Zyskind, O. Nathan, and A. Pentland, "Enigma: Decentralized computation platform with guaranteed privacy," *arXiv preprint arXiv:1506.03471*, 2015.

[25] P. Maymounkov and D. Mazieres, "Kademlia: A peer-to-peer information system based on the xor metric," in *International Workshop on Peer-to-Peer Systems*. Springer, 2002, pp. 53–65.