Data Management in Supply Chain Using Blockchain: Challenges and A Case Study

Hanqing Wu*, Jiannong Cao*, Yanni Yang*, Cheung Leong Tung*, Shan Jiang*, Bin Tang*,
Yang Liu†, Xiaoqing Wang†, Yuming Deng†
*The Hong Kong Polytechnic University, Hong Kong, China
†Supply Chain Platform of Alibaba Group, Alibaba Group Holding Limited, Hang Zhou, China
{cshwu,csjcao,csynyang,cscltung,cssjiang,csbint}@comp.polyu.edu.hk, {lionel.ly, robin.wxq, yuming.dym}@alibaba-inc.com

***Abstract*—Supply chain management (SCM) is fundamental for gaining financial, environmental and social benefits in the supply chain industry. However, traditional SCM mechanisms usually suffer from a wide scope of issues such as lack of information sharing, long delays for data retrieval, and unreliability in product tracing. Recent advances in blockchain technology show great potential to tackle these issues due to its salient features including immutability, transparency, and decentralization. Although there are some proof-of-concept studies and surveys on blockchain-based SCM from the perspective of logistics, the underlying technical challenges are not clearly identified. In this paper, we provide a comprehensive analysis of potential opportunities, new requirements, and principles of designing blockchain-based SCM systems. We summarize and discuss four crucial technical challenges in terms of scalability, throughput, access control, data retrieval and review the promising solutions. Finally, a case study of designing blockchain-based food traceability system is reported to provide more insights on how to tackle these technical challenges in practice.**

***Index Terms*—Supply Chain Management, Blockchain, Distributed Ledgers, Consensus**

I. Introduction

The global supply chain market surged over $13 billion in 2017 and is expected to soar past $19 billion by 2021 with the additional revenue opportunity from Software as a Service (SaaS) [1]. Although the supply chain industry has great potential for development, it suffers from a wide gamut of issues in supply chain management (SCM). To name a few, the lack of transparency and information sharing to delays in data retrieval affecting every stage of a logistics network. Furthermore, due to the centralized and separated systems in current SCM, product authentication and traceability cannot be achieved decently which the industry is struggling to handle.

The revolution of SCM relies on reliable and efficient data management, in which the data collected from supply chains are supposed to be stored, integrated, and retrieved with reliability and high efficiency. Facing the aforementioned issues, people are heading towards the application of the blockchain technology on SCM. Blockchain, originated from Bitcoin, is an implementation of an append-only ledger. It stores fully traceable and immutable records, which can be transformed from the data along with the supply chain, e.g., product and sales information. As such, blockchain enables the authenticity of the digitalized data in the supply chain. Meanwhile, it can be used as an overall system to integrate the data flow along with each step in the supply chain for efficient data management.

With much attention attracted from the blockchain for SCM, the researchers have conducted conceptual analysis [2][3][4] on the potential opportunities, advantages, and concerns when applying blockchain in different supply chains. From the perspective of logistics, blockchain has been positively recognized by the community. There have been some specific systems developed for some particular supply chain applications, for example, luxury and food supply chain [5][6][7]. Despite these great efforts, there are few studies focusing on the technical challenges for applying blockchain for SCM in practice.

In this paper, we put more emphasis on figuring out the technical challenges in blockchain regarding its application in SCM. In particular, the large number of stakeholders leads to scalability issue of the blockchain network. In terms of the huge amount of data generated from the supply chain, the throughput of the overall system and the latency of every single transaction should be guaranteed to make the system more user-friendly. Next, we discuss the issues about the access control in the blockchain system to prevent some data from being exposed to the competitors. Finally, the efficiency and reliability of data retrieval to trace the history information in the supply chain are investigated.

After identifying the four technical challenges, we present a case study about the food supply chain with the development of a blockchain-based food tracing system. The case study is significant because food safety is a major concern for the whole society. The system is built upon the Hyperledger Fabric for permissioned blockchain system. Functions, including the user access control, food data submission, data query and search are realized with the smart contract. Based on the developed system, we test the performance with respect the throughput for data submission and the speed of data query, so that we can discuss the effects that are key to the system performance when designing the blockchain.

In summary, the contribution of this paper lies in three aspects as follows

- We provide the insight of the potential opportunities to apply blockchain technology in SCM and present an exhaustive survey of existing blockchain-based SCM systems.

- We summarize the pain spots of existing SCM systems and four technical challenges in the design of blockchain SCM systems in practice.
- We implement a food tracing system based on permissioned blockchain for the food supply chain scenario.

The rest of this paper is organized as follows. Section II introduces the background of supply chain management and blockchain with the discussion on the application of blockchain for SCM. Section III presents the existing works on the investigation and studies of using blockchain for SCM. Section IV raises the technical challenges in designing the blockchain for fulfilling the requirements in SCM. The illustration of the food tracing system as a case study is given in Section V. Finally, Section VI summarizes the paper.

II. Supply Chain Management with Blockchain

In this section, we first introduce the background knowledge of supply chain and its management in brief. Then, the problems in SCM and the potential opportunities for blockchain to overcome the problems are discussed.

A. Briefing on SCM

Supply Chain (SC) is a system of all the activities and associated information flows involved in moving products or services from the supplier to the customer [8]. Supply chain management involves the management of the activities and information related to the sourcing, procurement, conversion, and all logistics [9]. Supply chain management can bring many financial, environmental and social benefits, e.g., improved resource utilization, the reduced cycle time from order to delivery, and early problem detection.

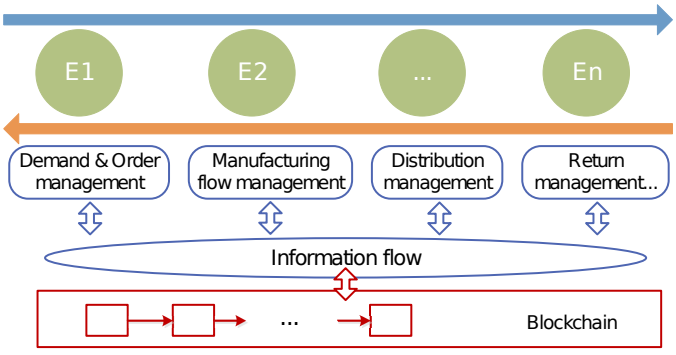

Fig. 1. SCM with Blockchain: E1, E2, ..., En represent the elements in supply chain, for example, the supplier and distributor

One of the key issues to realize the SCM functions, including demand and order management, manufacturing management distribution management and so on, is the data management [10]. The types of data in the supply chain involve but not limited to the inventory information, marketing information, and customer feedback. The circulation of those data within the supply chain is the core of its process and management.In Fig. 1, E1, E2, and En denote the stakeholders in the supply chain, e.g., supplier and distributor. For different management functions of the whole process, the information flow provides essential inputs. However, data management in current SC is occlusive and inefficient. In particular, some stakeholders store the data in a stand-alone and off-line way. Moreover, the information is exchanged via the postal system. A more effective way is to use the electronic data interchange (EDI) for the automation of the information flow. However, stand-alone and centralized data management brings some problems. First, data can be tampered which could intrude the data authenticity. Besides, system security can be violated. Finally, the retrieval of the data is time-consuming. Therefore, to promote the supply chain data management, we consider the following guidelines: (1) improve the coordination and information sharing within SC; (2) protect the data authenticity; (3) speed up the data retrieval.

B. Blockchain for SCM

The blockchain is an append-only list of blocks, each of which includes multiple pieces of data, managed by a peer-to-peer network adhering to a protocol for inter-node communication. The magic of blockchain lies in the protocol of validating new blocks, in short, consensus mechanism. The majority of the nodes will agree on the presence of each block via consensus algorithm after validating the data in the block. It can be quite difficult to tamper the data on the blockchain since most of the nodes will not admit it. There are various kinds of blockchain systems, including permissionless and permissioned blockchains, targeting for different application scenarios.

The nature of the blockchain technology brings about the features of system decentralization at the same time with the data immutability. These features provide potential opportunities for fulfilling the requirements of supply chain data management. First, by using the blockchain to store and manage the data flow in the supply chain, the information cannot be easily tampered and treated as reliable proof of existence. Second, data from different stakeholders in the supply chain can be integrated into the blockchain system rather than separately stored in individual systems, which not only helps for the data sharing but save cost and time for data retrieval. In the next section, we will introduce the existing works for applying blockchain for SCM.

III. Existing Works on SCM with Blockchain

There are some existing works performing conceptual studies on how to improve SCM by applying the blockchain technology. The adoption of blockchain for SCM is commonly related to supply chain data management. As shown in Fig. 1, the blockchain systems is connected to the information flow. The data for SCM is the input to the blockchain system. For data management, especially for the supply chain, the first issue is to collect the data where the sensing technologies can play their roles[11][12]. There are many works introducing the usage of IoT devices, e.g., RFID [13] and NFC [14], to collect the data from the physical world and convert it into digital information.

Second, after data collection, researchers try to understand and identify the application of blockchain in SCM via different approaches. Some works collect the feedbacks from the people in the community of logistics and economics by using questionnaires [4][13] so that they can gain more insights and the requirements of SCM from the industry. Some discuss the effects of blockchain in SCM based on different theories and framework, like Attributed of Innovation Framework [3] and Unified Theory of Acceptance and use of technology [15]. In general, the essential property of the blockchain, which is the decentralization, brings out the key features that are appealing to the SCM, including trustlessness, security, and authentication.Then, we can help to deal with the issues in SCM, e.g., traceability [2], cost-efficiency [4], and transparency [16].

Apart from the systematic analysis of the usage of blockchain in SCM, there are many deployed and conceptual systems for various specific SC scenarios, among which food supply chain is the hottest topic. The food provenance and safety issues are important problems to take care of [17]. There are some on-going projects for the food supply chain or food traceability, e.g., FarmShare [18], Provenance [19] and ripe.io [20]. More particular applications are about the wine [2] and agricultural food [21]. Pharma and drug industry also attract the help from blockchain [7][11] since healthcare is also an important social problem. There are also blockchain's applications in the post SC, sand SC, and excipient SC.

However, current studies, especially for those from the multiple disciplinaries, are mostly carried out from the perspective of the logistics, economics, and management. The advantages of importing the blockchain are appreciated by them, but they also argue that barriers are still present if the blockchain is put into practical use for SCM in terms of the immaturity of the current blockchain technology and the cost of its deployment. Meanwhile, only a few works consider the problems, for instance, security and privacy issues [5][22] in the technical design of the blockchain system. There are few studies focusing on how to make the blockchain system satisfy the requirements of SCM in practice. To this end, our paper put more emphasis on figuring out the technical challenges in Blockchain regarding its application in SCM. At the same time, we provide some experience in the design of the blockchain for the food supply chain as a representative case study.

IV. Technical challenges in blockchain for SCM

Supply chains typically raise various issues that are highly dependent on freight failure, human error, intended fraud, and others. However, when applying blockchain technology to the SCM, there will be more factors that will impact the system. These factors pose significant challenges in designing and implementing such systems. In the following, we present and analyze fours challenges of blockchain technology that affect, explicitly or implicitly, the supply chain.

A. Scalability

A large number of stakeholders are participating in modern supply chains, which are of the global range, together with a massive flow of newly created and time-sensitive information. All the data is poorly handled by modern SCM systems since there are few shared common database for stakeholders along with the supply chain to use. Blockchain can contribute significantly by providing a networked and decentralized database in order for all supply chain parties to join. However, the blockchain works in a decentralized fashion with stakeholders alongside the supply chain participating and interacting with each other, differs from the traditional SCM Electronic Data Interchange (EDI) systems which work in a centralized way with the system admin controlling the read and write access to the data. How the system can scale and operate with the increasing number of stakeholders and a large amount of generated transactional data is the prime challenge.

Network Scalability Blockchain platforms possess a specific mechanism that ensures data immutability along with the ledger called consensus. The scope of the consensus is to keep a general agreement between the nodes of the network about all submitted transactions. Such transaction information can be the timestamp, thus the order they occurred, the addresses of the sender and receiver, the amount transacted, the tags or electronic seal that accompany the ware and others. The essence of blockchain technology innovation and uniqueness originates from the use of consensus mechanisms. Today, blockchain platforms support different types of general agreement tools depending on the ledger level of access. A ledger can be public or private, and typical consensus algorithms constitute PoW (Proof-of-Work), PoS (Proof-of-stake), PBFT (Practical Byzantine Fault Tolerance), PoET (Proof of Elapsed Time), and PoA (Proof of Authority), respectively. Apparently, there is no common tool on modern supply chains that organizes and secures each step of the product. Hence errors, fraud, and ware failure are possible. Consensus algorithms constitute the core of the blockchain trust and node general agreement, with well-known methods known as "mining". On the contrary, such techniques hide dangers, such as 51% attack and reveal problems such as selfish mining which already happened in real life.

Storage Scalability With the number of transactions increasing day by day, the blockchain becomes heavy since all transactions have to be stored for validating the transaction. Currently, Bitcoin and Ethereum blockchains have exceeded 230GB [23] and 208GB [24] storage respectively. Large block size can increase the system throughput temporarily. However, the increased block size slows the propagation speed down, which leads to blockchain forks. So scalability is quite a tough problem in the blockchain. To solve the bulky blockchain problem, a novel cryptocurrency scheme was proposed in [25]. In the new scheme, old transaction records are removed by the network and a database named account tree is used to hold the balance of all non-empty addresses. In this way, nodes do not need to store all transactions to check whether a transaction is

valid or not. Besides lightweight client could also help fix this problem. A novel scheme named VerSum [26] was proposed to provide another way of allowing lightweight clients to exist. It allows lightweight clients to outsource expensive computations over large inputs. It ensures that the computation result is correct by comparing results from multiple servers.

B. Throughput

Various actions and procedures occur during the journey of a product inside the supply chain. They are often prone to human errors or even fraud or ware failure, which as a result, diminish system performance. With blockchain, the majority of activities can be represented as electronic transactions submitted on the ledger. In that case, those activities execute faster and without errors increasing SCM system performance. However, it is not easy to guarantee the system throughput in the blockchain.

While previous work has identified additional metrics, system throughput is the bottleneck issue and more challenging to address from a research perspective. Bitcoin's transaction throughput is a function of its block size and inter-block interval. With its current block size of 1 MB and 10 minute inter-block interval, the maximum throughput is capped at about seven transactions per second [27]; and a client that creates a transaction has to wait for at least 10 minutes on average to be sure that the transaction is included in the blockchain. In contrast, mainstream payment-processing companies like Visa confirm transactions within a few seconds and have a high throughput of up to 24,000 transactions per second.

Current research is focused on developing solutions to significantly improve blockchain performance while retaining its decentralized nature. Reparametrization of Bitcoin's block size and inter-block interval can improve performance to a limited extent estimated by a recent study at 27 transactions per second and 12 seconds, respectively. However, significant improvement in performance requires a fundamental redesign of the blockchain paradigm.

C. Fine-grained Access Control

The modern supply chains suffer from participants' societal fears about loss of privacy and data protection where any kind of data is available and can possibly be tampered. This makes many large companies unwilling to share their data which creates data silos along the supply chain. Blockchains offer a powerful contribution when it comes to those issues. The blockchain ledger not only contains immutable data but at the same time, participants' privacy can be highly respected and preserved with corresponding access control measures.

We classify blockchain data into two categories: the user identity and transactional data. For the user identity, permissionless blockchains, e.g., BTC, ETH, offers pseudonymity to its users in the sense that users only make transactions with newly generated addresses rather than real identities to avoid identity exposure. Thus, there is no longer any central party keeping users' private information. While inside a permissioned blockchain, total anonymity can be assured that participants are joining in an anonymized way while being authenticated in advance by an off-chain system, e.g., the government, FDA, of the supply chain. In this way, the supply chain system functions properly with participants real identity kept safely from the other users of the network, and it is guaranteed that they are legal participants. Nevertheless, in [28], the authors presented a method to link user pseudonyms to IP addresses even when users are behind NAT or firewalls. Moreover, each client can be uniquely identified by a set of nodes it connects to which can be learned and used to find the origin of a transaction.

However, the more valuable assets in the supply chain are the transactional data. Under some circumstances, participants would prefer to reveal their identities while the transactional data, e.g., manufacturing logs, retailer and consumer sales information, needs to be protected with fine-grained access control. Transactions in a permissionless blockchain are visible to the public while the read permission depends on the permissioned blockchain. The supply chain consortium could decide whether the stored data is public or restricted with levels of access control. Some works have been addressing this issue in the past few years. In [29], the authors propose a decentralized personal data management system, implemented on the blockchain, that ensures the user ownership of their data. In [30], it proposed a new approach based on blockchain to publish the policies expressing the right to access a resource and to allow the distributed transfer of such right among users. In [31], a blockchain-based platform for healthcare information exchange is proposed to satisfy the requirements of both privacy and authenticity.

D. Data Retrieval

When adopting blockchain technology with SCM, sharing accurate and timely information throughout a supply chain yields significant benefits to all participants along the SCM. Every recordable data requires peer-to-peer verification, which can be a time-consuming work with the number of blocks involved when tracing the linkage of data backward. With the data pass through the peer-to-peer verification and the block containing the data is appended to the blockchain, it is important to retrieval the desired data from the blockchain in an efficiency and reliability way.

Participants have different requirements and expectation for data retrieval. Wholesalers want to trace forward the product to get the accurate situation of sales in order to make a better marketing strategy and increase the revenue in return. The consumers want to know the authenticity of the product quickly so that they can confidently decide the purchase without too much hesitation. Those requirements place challenges in data retrieval. The efficiency of data retrieval means that the queried data should return the results within a reasonable time range. The reliability of data retrieval means that the return results should not be incomplete and tampered.

Retrieval by full node In existing mainstream blockchain applications, it is not easy to achieve efficient or reliable

data retrieval. For example, in both BTC and ETH, a user who wants to do the query has to be a full node that fully downloads every block and transaction and check them against blockchain's consensus rules which may take days for the startup. This poses a heavy burden on both data storage and network bandwidth, and it is user unfriendly. After that, the blockchain data is locally available and can be imported into designated databases to rebuild indexes for the retrieval. However, it is obvious that this local data retrieval manner is not efficient or even practical for nodes with low-end hardware.

Retrieval by lightweight node Online data retrieval is possible because a node can also choose to be a lightweight node which does not download the complete blockchain. Instead, the light nodes download the block headers only to validate the authenticity of the transactions. Because of this reason light nodes are easy to maintain and run. However, light nodes need full nodes to connect to the network and therefore, the effectiveness of data retrieval completely dependent on the full nodes to function. More downsides come from privacy and security consideration. For the privacy, the light nodes typically send transactions to a trusted third party which allows the trusted third party to spy on all the users past and future activities. For the security, the light nodes may skip several security steps which can leave the user vulnerable, and the trusted third party can possibly be a malicious node to launch the middleman attack.

Besides data storage, the blockchain is also supposed to provide data retrieval service to the stakeholders along the supply chain. However, it is non-trivial to search for data on blockchain efficiently while preserving data privacy. In particular, the data stored on blockchain are typically encrypted, which hinders the development of efficient search algorithms. Moreover, the search operations are special in SCM, for example, forward search and backward search.

V. Case Study

In this section, we first introduce the system framework to build a blockchain-based SCM system. Then, we demonstrate the implementation based on Hyperledger Fabric with the functionality design, system architecture, and implementation details. Finally, we conduct extensive experiments to analyze the system performance in terms of response times of user registration, data submission, and data query.

A. System Overview

Food safety is the major concern for the society nowadays. Problems including food fraud, illegal production and food-borne illness in the food supply chain have resulted in much damage to customers' health and loss in the food industry. In fact, governments and many organizations have paid close attention to the food safety problems and taken measures to deal with it. However, there is a long and tough way to go.

Our system adopts the concept of Blockchain as a Service (BaaS). BaaS is an offering that allows customers to leverage cloud-based solutions to build, host and use their own blockchain apps and functions. With the help of BaaS, clients only need to concern about the functions they want to realize. From the technical view, blockchain service providers provide flexible choices about different components in the data layer, consensus layer or smart contract layer in the blockchain infrastructure. In addition, it could be much easier to operate multiple chains simultaneously.

Therefore, our blockchain-based food traceability system is built upon the Hyperledger Fabric, which is an open-source and permissioned distributed ledger technology platform. The reason why we use the Fabric is that it provides modular and configurable architecture, which is potential for the realization of BaaS. Besides, it is permissioned which is similar to the federated blockchain with moderate-trust among different parties and suitable for the supply chain applications.

The general framework of our food traceability system is shown in Fig. 2. The bottom layers are network layer and data layer. The submitted data are aggregated into blocks and serialized into the form of a chain. When preparing each data block, cryptographic functions can be used for data representation and connecting consistent blocks. The proposed data and blocks are broadcasted via different protocols. The second layer is the consensus layer, including leader election and transaction packing. There are many consensus algorithms, e.g., PoW and PBFT. In the Fabric, we can insert our own designed consensus algorithms. Currently, we use the PBFT for leader election in the system. Then, the selected node would select and pack data pieces into one block. The third layer is the contract layer, many functions can be realized in the smart contract. The upper layer is the application layer, of which the characteristics are subject to the purposes of different applications.

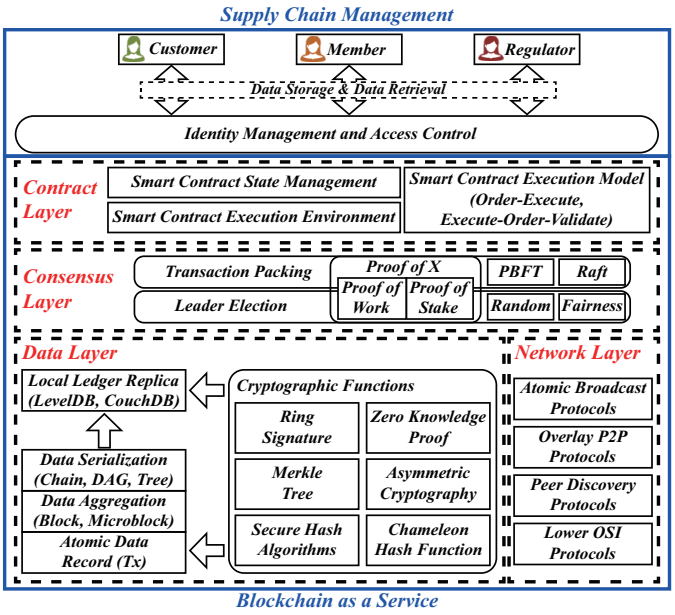

Fig. 2. Framework of the blockchain-based food traceability system

In our system, there are three kinds of identities, which

are the regulator, member, and customer. Regulators have the highest level of access to the blockchain, all the information can be visible to regulators for censorship. In practice, the regulator can be food safety association or other official organizations. Members refer to the suppliers, companies and other parties in the food supply chain. They have rights to write in, i.e., submitting information to the blockchain, and only the "one back, one up" food information is visible to them. This means that they can only search the last source and next step of the food transferred in the supply chain. Customers do not need to write in, and the information exposed to them only tells the information about the main producer of the food product and the origin areas of the raw food. The intermedia information will be hidden for the customers. The access control of the three identities is realized by recording their identity information in the blockchain. Their own private key is used for log-in. The identify checking is implemented by the smart contract.

The functionality of the system is briefed in Fig. 3. Regulators and federated members are required to log into the system before further operation. New federated members should register for legitimate access to the system, which would be given by the regulators. The registration records and access records would all be stored on the blockchain as proof. Federated members can submit food information and query previously submitted data by invoking smart contract. The submitted food information will be verified by the peers and published on the blockchain. For the regulators, they are search any information existed in the ledger. While for customers, they only can query the information about the food they bought from retailers. There are different smart contracts from implementing the query and search request sent by the regulators, members and, customers.

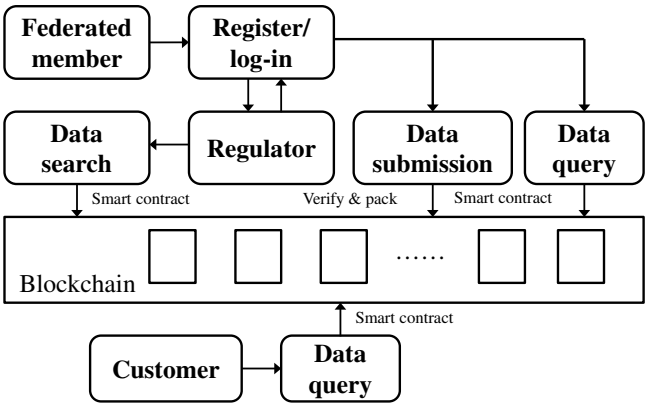

Fig. 3. Functionality of the food traceability system

For setting the nodes for the food blockchain, there should be full nodes and light nodes in the network. The full nodes can be set by the regulators, while the light nodes are created by the members. One reason for doing this is that only the regulators could see all the information across the food supply chain due to the privacy concerns of business competitors. Another reason is that, if only regulators have the full nodes, the information can be trustless for the members if regulators tamper the original data. Therefore, a group of members could have light nodes that store their own food information.

B. System Implementation

To demonstrate the effectiveness and practicability of the proposed system from case study, we implement the food traceability system in a minimal-viable-product version. The system is developed under the framework of Hyperledger Fabric.

The currently system is developed under Hyperledger Fabric v1.1 [32]. The system is implemented on our virtual machines with Docker [33]. Each virtual machine consists of 1 core 2 threads of Intel Core i7-8809G 4.2Ghz CPU with 4GB of DDR4 DRAM and 40GB of NVMe SSD, running on Ubuntu 16.04.5.

For the *System and Data Setup*, we deploy 4 nodes within department intranet to evaluate the system. The node is capable of running 4 docker containers for different roles, i.e., 2 for client peer, 1 for orderer and 1 for endorser, at the same time. The responsibility of network roles are explained as followed. Each node has individual IP such that they can communicate with each other.

- Client: a client that submits an actual transaction-invocation to the endorsers, and broadcasts transaction-proposals to the ordering service.
- Orderers: a node that commits transactions and maintains the state and a copy of the ledger
- Endorser: a node running the communication service that implements a delivery guarantee, such as atomic or total order broadcast.

C. System Analysis

In the *System Performance Test*, for the function of *Member Registration*, we invoke the chaincode (smart contract) with 10, 30, and 50 concurrent jobs at the same time and repeat the experiment 10 times to get the average time for the registration which is successfully finished and recorded on the blockchain, as depicted in Fig. 4 as cumulative distribution function (CDF) plot. We also apply the same testing parameters to the function of *Transaction Uploading* and *Data Retrieving* as depicted in Fig. 5 and Fig. 6 as CDF plots respectively. For the *data uploading*, including system registration and transaction uploading, these two functions will invoke both 1436 individual jobs. For the *data researching*, the total number of the query is 1000.

With the CDF plots from the experimental data, it is no hard to find that with the increased number of concurrent jobs within one chaincode, the system response time will increase linearly. Still, the time of data retrieving will increase significantly with the increase of concurrent jobs. When we send 10 queries to the blockchain, we can get the most results within 0.3 second. However, the average response time of 50 queries takes close to 1 second. The reason is that the system registration and transaction uploading are treated as writing

operations to the blockchain while data retrieving is treated as reading operations. During the query, it may involve multiple transactions' retrieving and validation since these transactions are linked in the blockchain.

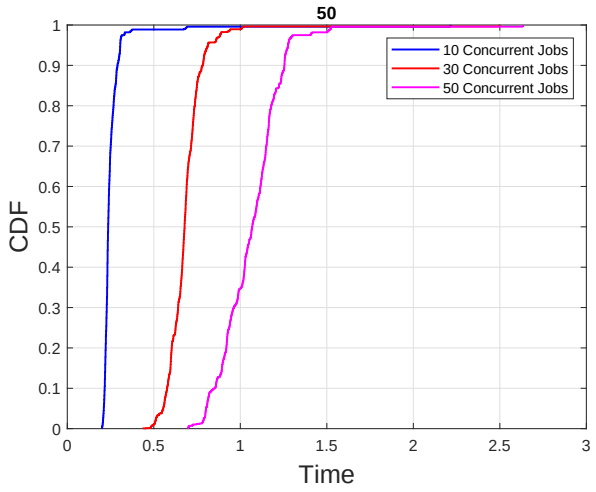

Fig. 4. CDF of Member Registration Time

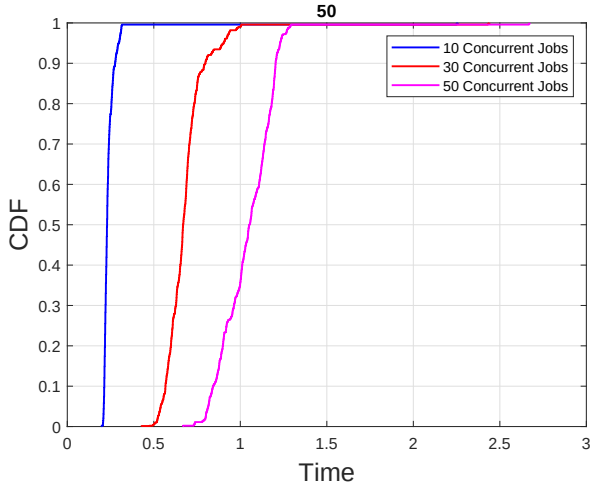

Fig. 5. CDF of Transaction Uploading Time

1) Throughput Analysis: The current system TPS is not high due to the following factors: 1) The current Hyperledger Fabric framework is a generic framework supporting different applications, and it is not optimized for high TPS; 2) The transaction process flows require that the system needs to process and verify transactions one by one instead of parallel; 3) The transaction size is big and not optimized yet.

In particular, we analyze the *Transaction Size* as major impact factor to the system. When using naive design, i.e., one food item submission in each transaction, the system throughput, write speed, is low. The possible solutions can be: 1) bundling multiple item submission in the transaction; 2) increase the number of transactions per block. We can

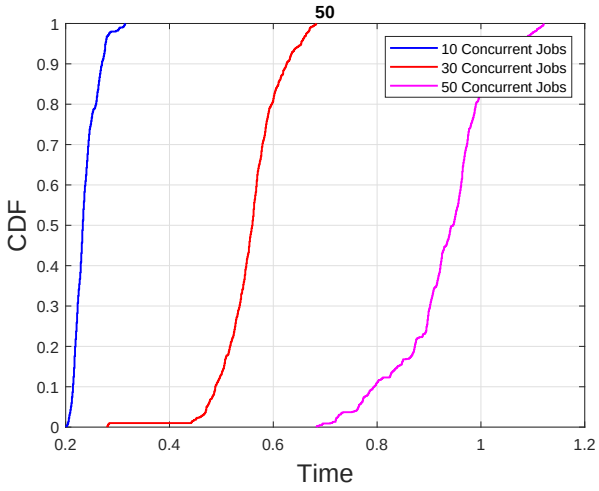

Fig. 6. CDF of Data Retrieving Time

also achieve higher utility rate (payload/total block size) to increase the efficiency by packing related food items together, in a single transaction atomically. Moreover, more complex packing algorithm, with the higher computational requirement, is needed to lower the real-time capacity. Overall, further study is needed to propose a framework on adjusting the block size and number of transactions to balance system throughput and real-time capability.

2) Privacy Analysis: In the current system, we focus the privacy protection with following safety precautions: 1) All the data, e.g., transactions, logs, system events, on the blockchain are encrypted with private-public key pairs. Only the user with the corresponding private key, in the form of owning or being given access rights, can decrypt and view the data; 2) The transaction data can only be seen from neighbor hops, by system default configuration, when doing the data retrieval. This guarantees the traceability of the system while safe-guarding the privacy of data. Only the administrator, e.g., the government, FDA, can trace the entire history when consumer sends an appealing request about the disputed product.

VI. Conclusion

In this paper, we have introduced the blockchain technology in supply chain data management. In particular, we looked into the potential opportunities of applying blockchain for SCM and summarizing the existing works on blockchain for SCM. We studied the requirements from SCM when adopting blockchain technology and also demonstrated the key technical challenges in the design of the blockchain for meeting the demands of the supply chain in practice. A case study is presented to address the aforementioned issues and introduces our proposed food safety tracing system. We implemented this system based on the permissioned supply chain for the food supply chain scenario.

For the future work, we will consider integrating real-world supply chain data with the current system and deploy the system on more federated nodes.

ACKNOWLEDGMENT

This work is supported by Alibaba Innovative Research (AIR) Program by Alibaba (China) Co., Ltd. - H-ZG6N, Hong Kong RGC Research Impact Fund (RIF) - R5034-18, and Hong Kong RGC Collaborative Research Fund - CityU C1008-16G.

REFERENCES

[1] "Gartner report on supply chain management," https://www.gartner.com/newsroom/id/3747517, Accessed: 2019-04-15.

[2] K. Biswas, V. Muthukkumarasamy, and W. Lum, "Blockchain based wine supply chain traceability system," in *Future Technologies Conference*, Nov. 2017.

[3] M. Dobrovnik, D. M. Herold, E. W. M. Frst, and S. Kummer, "Blockchain for and in logistics: What to adopt and where to start," *Logistics*, vol. 2, no. 3, 2018.

[4] N. Hackius and M. Petersen, "Blockchain in logistics and supply chain : trick or treat?" *Proceedings of the Hamburg International Conference of Logistics (HICL)*, pp. 3–18, 2017.

[5] L. Xu, L. Chen, Z. Gao, Y. Lu, and W. Shi, "Coc: Secure supply chain management system based on public ledger," in *26th International Conference on Computer Communication and Networks, ICCCN 2017, Vancouver, BC, Canada, July 31 - Aug. 3, 2017*, 2017, pp. 1–6.

[6] K. Korpela, J. Hallikas, and T. Dahlberg, "Digital supply chain transformation toward blockchain integration," in *50th Hawaii International Conference on System Sciences, HICSS 2017, Hilton Waikoloa Village, Hawaii, USA, January 4-7, 2017*, 2017, pp. 1–10.

[7] J.-H. Tseng, Y.-C. Liao, B. Chong, and S.-w. Liao, "Governance on the drug supply chain via gcoin blockchain," *International Journal of Environmental Research and Public Health*, vol. 15, pp. 1055–1063, 2018.

[8] "Wikipedia supply chain," https://en.wikipedia.org/wiki/Supply_chain, Accessed: 2019-04-15.

[9] "Inc., advanced solutions international," https://cscmp.org/CSCMP/Educate/SCM_Definitions_and_Glossary_of_Terms/CSCMP/Educate/SCM_Definitions_and_Glossary_of_Terms.aspx, Accessed: 2019-04-15.

[10] M. A. Waller and S. E. Fawcett, "Data science, predictive analytics, and big data: a revolution that will transform supply chain design and management," *Journal of Business Logistics*, vol. 34, no. 2, pp. 77–84, 2013.

[11] T. Bocek, B. B. Rodrigues, T. Strasser, and B. Stiller, "Blockchains everywhere - a use-case of blockchains in the pharma supply-chain," in *2017 IFIP/IEEE Symposium on Integrated Network and Service Management (IM), Lisbon, Portugal, May 8-12, 2017*, 2017, pp. 772–777.

[12] N. Kshetri, "1 blockchain's roles in meeting key supply chain management objectives," *Int J. Information Management*, vol. 39, pp. 80–89, 2018.

[13] K. Toyoda, P. T. Mathiopoulos, I. Sasase, and T. Ohtsuki, "A novel blockchain-based product ownership management system (POMS) for anti-counterfeits in the post supply chain," *IEEE Access*, vol. 5, pp. 17 465–17 477, 2017.

[14] N. Alzahrani and N. Bulusu, "Block-supply chain: A new anti-counterfeiting supply chain using NFC and blockchain," in *Proceedings of the 1st Workshop on Cryptocurrencies and Blockchains for Distributed Systems*, ser. CryBlock'18, 2018, pp. 30–35.

[15] K. Francisco and R. Swanson, "The supply chain has no clothes: Technology adoption of blockchain for supply chain transparency," *Logistics*, vol. 2, p. 2, 01 2018.

[16] J. Biggs, S. R. Hinish, M. A. Natale, and M. Patronick, "Blockchain: Revolutionizing the global supply chain by building trust and transparency," *Rutgers University, New Jersey*, 2018.

[17] D. Tse, B. Zhang, Y. Yang, C. Cheng, and H. Mu, "Blockchain application in food supply information security," in *2017 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM)*, Dec 2017, pp. 1357–1361.

[18] "Farmshare," http://farmshare.us, Accessed: 2019-04-15.

[19] "Provenance," https://www.provenance.org, Accessed: 2019-04-15.

[20] "Ripe.io," https://www.ripe.io/, Accessed: 2019-04-15.

[21] F. Tian, "A supply chain traceability system for food safety based on haccp, blockchain & internet of things," in *2017 International Conference on Service Systems and Service Management*. IEEE, 2017, pp. 1–6.

[22] K. Leng, Y. Bi, L. Jing, H. Fu, and I. V. Nieuwenhuyse, "Research on agricultural supply chain system with double chain architecture based on blockchain technology," *Future Generation Comp. Syst.*, vol. 86, pp. 641–649, 2018.

[23] "Bitcoin Blockchain Size," https://www.blockchain.com/charts/blocks-size, Accessed: 2019-04-15.

[24] "Ethereum Blockchain Size," https://etherscan.io/chartsync/chaindefault, Accessed: 2019-04-15.

[25] "The mini-blockchain scheme," https://cryptonite.info/files/mbc-scheme-rev3.pdf, Accessed: 2019-04-15.

[26] J. van den Hooff, M. F. Kaashoek, and N. Zeldovich, "Versum: Verifiable computations over large public logs," in *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security*, ser. CCS '14. New York, NY, USA: ACM, 2014, pp. 1304–1316.

[27] "Bitcoin Blokchain Transactions Per Second," https://www.blockchain.com/charts/n-transactions-per-block, Accessed: 2019-04-15.

[28] A. Biryukov, D. Khovratovich, and I. Pustogarov, "Deanonymisation of clients in bitcoin P2P network," in *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security, Scottsdale, AZ, USA, November 3-7, 2014*, 2014, pp. 15–29.

[29] G. Zyskind, O. Nathan, and A. Pentland, "Decentralizing privacy: Using blockchain to protect personal data," in *2015 IEEE Security and Privacy Workshops (SPW)*. Los Alamitos, CA, USA: IEEE Computer Society, may 2015, pp. 180–184.

[30] D. D. F. Maesa, P. Mori, and L. Ricci, "Blockchain based access control," in *Distributed Applications and Interoperable Systems - 17th IFIP WG 6.1 International Conference, DAIS 2017*, 2017, pp. 206–220.

[31] S. Jiang, J. Cao, H. Wu, Y. Yang, M. Ma, and J. He, "Blochie: A blockchain-based platform for healthcare information exchange," in *2018 IEEE International Conference on Smart Computing (SMARTCOMP)*, June 2018, pp. 49–56.

[32] "Hyperledger Fabric," https://hyperledger-fabric.readthedocs.io/en/release-1.1, Accessed: 2019-04-15.

[33] "Docker," https://www.docker.com, Accessed: 2019-04-15.