Research Article

Light Repository Blockchain System with Multisecret Sharing for Industrial Big Data

Hefeng Chen,1 Hsiao-Ling Wu,2 Chin-Chen Chang3, and Long-Sheng Chen2

1Computer Engineering College, Jimei University, Xiamen 361021, China
2Department of Information Management, Chaoyang University of Technology, Taichung 413, Taiwan
3Department of Information Engineering and Computer Science, Feng Chia University, Taichung 407, Taiwan

Correspondence should be addressed to Chin-Chen Chang; alan3c@gmail.com

Received 1 March 2019; Accepted 4 August 2019; Published 16 October 2019

Guest Editor: Fagen Li

Copyright © 2019 Hefeng Chen et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Blockchain is an emerging technology that promises many exciting applications in various fields, including financial, medical, energy, and logistics management. However, there are still some limitations in the existing blockchain framework that prevents its widespread adoption in the commercial world. One important limitation is the storage requirement, wherein each blockchain node has to store a copy of the distributed ledger. Thus, as the number of transactions increases, this storage requirement grows quadratically, eventually limiting the scalability of a blockchain system. Moreover, the public ledger in a blockchain framework allows anyone in the network to audit the transaction, which may not be favourable in some privacy-sensitive applications. In this paper, a secret-sharing scheme is proposed to reduce the size of the blockchain transactions. Each transaction block is divided into \( t \) parts, and the size of each part is \( 1/t \) size of transaction block. We use the secret-sharing mechanism to share \( t \) parts into \( n \) shares. Hence, each node stores not one transaction but one share in the blockchain system. The proposed scheme can eventually reduce the storage cost of a blockchain transaction by \( 1/t \) without introducing an additional recovery communication cost; however, robustness is reduced in node failure as a tradeoff. Meanwhile, the proposed scheme was more efficient and secure compared to other state-of-the-art schemes that aim to reduce blockchain storage for industrial big data.

1. Introduction

Internet commerce has become relatively entirely dependent on financial institutions that are trusted third parties to handle electronic payments. Although the system is suitable for most transactions, it still has the inherent weaknesses in trust-based models. Blockchain is a paradigm-shifting technology that has emerged over the past decade, which is based on peer-to-peer communication technology, network theory, and cryptography [1]. In a traditional centralized database model, the central server stores every transaction record, which can be disastrous if the central server is being attacked. In contrast, blockchain technology is a distributed solution without any third-party trust problem. Since every node in the blockchain network keeps a copy of the public ledger, it is possible to audit the transactions locally without referring to a centralized authority. Therefore, blockchain can also be viewed as a distributed ledger system, which eliminates the disadvantage of a traditional centralized database model. However, the storage cost in blockchain is huge [2], which is one of the factors currently limiting the widespread adoption of blockchain technology. Since each node needs to store the public ledger locally, the storage cost of an entire blockchain network grows quadratically.

To resolve the storage issue in blockchain, Dai et al. [3] proposed a low-storage blockchain system that employs network coding theory to divide the transaction data into multiple blocks and then stores the blocks at different nodes. This can realize distributed storage (DS) by recovering the transaction data through network coding (NC). They proposed two kinds of NC-based DS, one is deterministic rate (NC-DRDS) and the other is rateless (NC-RLDS), to deal with a fixed and variable number of blockchain nodes, respectively. In another approach, Dorri et al. [4] proposed a memory-optimized and flexible blockchain (MOF-BC) that allows the user to summarize or remove part of the “aged”
2 Security and Communication Networks

Blockchain and Storage Issues

Blockchain is a distributed ledger system that requires each participating node to store a copy of ledger for the transaction records. Every time a transaction block is created, it is first verified by the neighbouring nodes and then goes through the consensus (mining) process by solving a difficult cryptographic puzzle. One of the nodes that successfully solve this puzzle is rewarded with some incentives. This transaction block is then added to the ledger, wherein each block is related to the previous block through a chain of hash values. This data structure suffers from scalability issue, as the storage room required to keep the entire ledger is growing quadratically when the number of blockchain nodes increases.

3 Overview of Shamir’s Secret-Sharing Scheme

3.1. Shamir’s (r, n)-Threshold Secret Sharing. In 1979, Shamir developed a (r, n)-threshold secret-sharing scheme [13] based on polynomial interpolation, wherein a univariate polynomial \( y = f(x) \) of degree \( r - 1 \) is uniquely defined by \( r \) points \( (x_j, y_j) \) with distinct \( x_j \), for \( j = 1, 2, \ldots, r \). The scheme can decompose one secret into \( n \) shares, with \( r \) shares required to recover the original secret, where \( r \leq n \). However, the secret cannot be recovered with less than \( r \) shadows. It consists of the following two phases.

3.2. Shadow Distribution Phase. The trusted dealer starts with a secret integer, \( S \geq 0 \), that is, to be distributed among \( n \) users. Thus, the dealer runs the following:

(1) Choose a prime \( p > \max(n, S) \).

(2) Randomly select \( r - 1 \) independent coefficients \( a_1, a_2, \ldots, a_{r-1} \in \mathbb{Z}_p \) to constitute a random polynomial with \( r - 1 \) degree over \( \mathbb{Z}_p \):

\[
f(x) = S + \sum_{j=1}^{r-1} a_j x^j \mod p.
\]

(3) Choose \( n \) distinct nonzero elements of \( \mathbb{Z}_p \), denoted as \( x_i \), for \( 1 \leq i \leq n \).

(4) Compute \( s_i = f(x_i) \mod p \), \( 1 \leq i \leq n \) and securely transfer the share \( s_i \) to user \( U_i \), along with the public index \( x_i \).

3.3. Secret Reconstruction Phase. Assume that \( r \) users, \( U_1, U_2, \ldots, U_r \), pool their shares to compute the secret \( S \). Their shadows provide \( r \) distinct points \( (x_j, s_j) \), \( 1 \leq j \leq r \), which allow the computation of the coefficients of \( f(x) \) by Lagrange interpolation. The secret, \( S \), can be expressed as

\[
S = f(0) = \sum_{j=1}^{r} s_j c_j \mod p, \quad (2)
\]

where \( c_j = \prod_{1 \leq k \leq r, k \neq j} (x_j / x_k - x_j) \mod p \), for \( 1 \leq j \leq r \).
4. Blockchain Transactions with Secret-Sharing Scheme

In this section, we describe our proposal of using secret-sharing scheme to reduce the storage cost of blockchain. In our sharing scheme, each block is stored in a distributed manner among all nodes. Consider the transaction block \( a \), our idea inspired by Parakh and Kak [20]. The flowchart of sharing process and reconstruction process is shown in Figure 2.

Firstly, we divide the block \( a \) into \( t \) pieces of size \([a]/t\), denoted as \( a_1, a_2, \ldots, a_t \), such that their concatenation \( a = a_1 || a_2 || \ldots || a_t \). All the arithmetic performed is finite field arithmetic on \( \mathbb{Z}_p \), where \( p \) is larger than \( a_i (1 \leq i \leq t) \) and \( n \). Then, we share them recursively as follows. The first step is to run Shamir’s (2, 2) secret sharing for the piece \( a_t \). That is, generate randomly a 1-degree polynomial \( g_1(x) \) whose constant term is \( a_t \), and compute \( g_{1,1} = g_1(1) \) and \( g_{1,2} = g_1(2) \) to obtain two shares for \( a_t \). Next, generate a 2-degree polynomial \( g_2(x) = a_2 + g_{1,1}x + g_{1,2}x^2 \mod p \) whose constant term is the next piece \( a_2 \), and the coefficients are the previous two shares. By induction, assume that we have obtained \( i \) shares of the piece \( a_{i-1} \) \((2 \leq i \leq t - 1)\), denoted as \( g_{i-1,1}, g_{i-1,2}, \ldots, g_{i-1,i} \), we can generate a \( i \)-degree sharing polynomial for the next piece \( a_i \) by using \( g_{i-1,j} \) as the \( j \)-th coefficient term for \( j = 1, 2, \ldots, i \). At the last step, we generate a \( t \)-degree sharing polynomial:

\[
g_t(x) = a_t + g_{t-1,1}x + g_{t-1,2}x^2 + \cdots + g_{t-1,t}x^{t-1} \mod p, \tag{3}
\]

and the final \( n \) shares are

\[
g_1(x_1), g_1(x_2), \ldots, g_t(x_n), \tag{4}
\]

where \( x_1, x_2, \ldots, x_n \) are \( n \) public indexes of nodes. The sharing scheme is shown in Algorithm 1.

The reconstruction is an inverse process that is carried out in a backward and first-out manner. Any \( t + 1 \) of the shares can interpolate a \( t \)-degree polynomial \( g_t(\cdot) \) with the constant term \( a_t \). Then, interpolate a \( t - 1 \)-degree polynomial \( g_{t-1}(\cdot) \) with the constant term \( a_{t-1} \), by taking the \( j \)-th coefficient of \( g_t(\cdot) \) as the point at \( x = j \), for \( j = 1, 2, \ldots, t \). The recursion repeats until \( a_1 \) is obtained. Then, the whole block is given by the concatenation \( a_1 || a_2 || \ldots || a_t = a \). Algorithm 2 shows how to reconstruct the shared transaction block.

In Figure 3, a block \( a \) is divided into 4 pieces \( a_1 = 19, a_2 = 27, a_3 = 4, \) and \( a_4 = 15 \). Assumed that \( a \) is to be shared between 7 parties such that any 5 of them can reconstruct all the 4 secrets. We can now use a prime \( p = 31 \) and random number \( r = 11 \).

\[\begin{array}{c}
\alpha \\
\alpha_1 \\
\alpha_2 \\
\alpha_3 \\
\alpha_4 \\
\end{array}\]

**Theorem 1.** Our proposed scheme divides the original file \( a \) into \( n \) pieces in such a way that (i) any \( t + 1 \) packets can recover the original file \( a \), but (ii) no group of \( t \) pieces can do so.

**Proof.** Based on the Lagrange interpolation, given \( t \) points \((u_{i_1}, y_{i_1}), (u_{i_2}, y_{i_2}), \ldots, (u_{i_t}, y_{i_t})\) with distinct \( u_{i_j} \)'s, there is a unique polynomial of degree \( t \) such that \( y_{i_j} = g(u_{i_j}) \), for \( 1 \leq j \leq t + 1 \), i.e.,

\[
g(x) = \sum_{j=1}^{t+1} y_{i_j} l_{i_j}(x) \mod p, \tag{5}
\]

where \( l_{i_j}(x) = \prod_{1 \leq k \leq t+1, k \neq i_j} (x - u_k) / (u_{i_j} - u_k) \mod p \).

Therefore, \( a_i = g(0) \) can be reconstructed correctly. To reconstruct the next piece \( a_{i-1} \), a \((t - 1)\)-degree polynomial is generated by taking the \( k \)-th coefficient of \( g(x) \) as the point at \( x = k \), for \( k = 1, 2, \ldots, t \). Correctness is still guaranteed by Lagrange interpolation. Then, based on recursion, \( a_t, a_{t-1}, \ldots, a_2, a_1 \) can be correctly reconstructed one by one.

On the contrary, if only \( t \) of these \( n \) packets are available, it is not suffice to compute \( a_i \)'s. For each candidate value \( a_i \in \{0, 1, \ldots, p - 1\} \), we can construct a unique polynomial \( g(x) \) of degree \( t \) such that \( g(0) = a_i \mod p \) and \( g(x_i) = y_{i} \mod p \) for \( 1 \leq j \leq t + 1 \). These \( p \) possible polynomials have equal probability; therefore, the real value of packets cannot be determined.

Moreover, we can improve the above proposed scheme to achieve the verifiable property. In general, each transaction block in the blockchain system (e.g., Bitcoin) consists of a block header and a data block. All transactions are included in the data block as leaf nodes of a Merkle tree [21], and the hash root of the Merkle tree is computed in the block header (see Figure 4). Denote the block header and data block of a transaction block, \( a \), as \( \beta \) and \( \delta \), respectively. Now, we perform the above sharing scheme for \( \beta \) and \( \delta \), independently. After reconstructing \( \beta \) and \( \delta \), the correctness of \( \delta \) can be verified by the root hash value in block header \( \beta \).
1. Sharing process

Transaction block $\alpha$

- $a_1$
- $a_2$
- $a_t$

$n$ shares $\{\beta_1, \beta_2, \ldots, \beta_n\}$

2. Reconstruction process

Transaction block $\alpha$

- $a_1$
- $a_2$
- $a_t$

$t + 1$ shares $\{\beta_{a_1}, \beta_{a_2}, \ldots, \beta_{a_t}\}$

$t + 1$ public indexes $\{x_{a_1}, x_{a_2}, \ldots, x_{a_t}\}$

**Algorithm 1:** Sharing process.

Given the block $\alpha$ and $n$ public indexes $\{x_1, x_2, \ldots, x_n\}$

- Divide $\alpha$ into $t$ equal length pieces $a_1, a_2, \ldots, a_t$
- Select a large prime $p > \max\{a_1, a_2, \ldots, a_t, n\}$
- Select a random number $r \in Z_p$, and construct a 1-degree polynomial $g_1(x) = \alpha_1 + rx \pmod{p}$
- Compute $g_{i+1} = g_i(1)$ and $g_{i+2} = g_i(2)$

for $i = 2$ to $t - 1$ do
- Construct a $i$-degree polynomial $g_i(x) = \alpha_i + g_{i-1}x + g_{i-2}x^2 + \cdots + g_{i-k}x^k \pmod{p}$

for $i = 1 + 1$ do
- Compute $g_i(1)$
end for

end for

Construct a $t$-degree polynomial $g_t(x) = \alpha_t + g_{t-1}x + g_{t-2}x^2 + \cdots + g_{t-k}x^k \pmod{p}$

for $i = 1$ to $n$ do
- Compute $\beta_i = g_i(x_i)$
end for

Distribute the share $\beta_i$ to the corresponding node with public index $x_i$, for $i = 1, \ldots, n$

**Algorithm 2:** Reconstruction process.

Given any $t + 1$ shares $\beta_{x_1}, \beta_{x_2}, \ldots, \beta_{x_n}$

Interpolate $t + 1$ points, $(x_1, \beta_{x_1}), (x_2, \beta_{x_2}), \ldots, (x_n, \beta_{x_n})$, to generate a $t$-degree polynomial $g_t(x) = \alpha_t + g_{t-1}x + g_{t-2}x^2 + \cdots + g_{t-k}x^k \pmod{p}$

Compute the constant term $\alpha_t = g_t(0)$ and the coefficients

for $i = k + 1$ to $1$ do
- Interpolate $i + 1$ points $(1, g_{1,1}), (2, g_{1,2}), \ldots, (i + 1, g_{1,i+1})$ to generate a $i$-degree polynomial $g_i(x) = \alpha_i + g_{i-1}x + g_{i-2}x^2 + \cdots + g_{i-k}x^k \pmod{p}$
- Compute $\alpha_i = g_i(0) \pmod{p}$
end for

return $\alpha = \alpha_1 + \alpha_2 + \cdots + \alpha_t$

5. Practical Implementation Issues

5.1. Storage Cost. Assuming that the block header is stored in $F_q$ and the data block is stored in $F_q$, then the blockchain’s storage cost per transaction per peer is

$$\log_2 \eta + \log_2 q.$$  \hspace{1cm} (6)

The storage cost of the DSB proposed by Raman and Varshney [5] is
\[
\log_2 \eta + 2 \log_2 q. \tag{7}
\]

This is effectively smaller compared to the traditional blockchain. Kim et al. [6] proposed the LSS scheme to further reduce the storage cost to
\[
\frac{\log_2 \eta}{t} + \log_2 q. \tag{8}
\]

The comparison of storage cost for different schemes is detailed in Table 1.

Compared to the traditional blockchain, our proposed scheme can effectively reduce the storage cost by \(1/t\), so the storage cost becomes \((\log_2 \eta + \log_2 q)/t\). From Table 1, it is clear that our proposed scheme is more efficient than other prior work including Kim et al.’s [6]. This can effectively reduce the deployment and maintenance cost in blockchain.

5.2. Recovery Communication Cost. Our proposed scheme can effectively reduce the storage cost in blockchain communication, but it also introduces some additional communication cost. In traditional blockchain, each node keeps a public ledger locally. If any node failure happens, the failed node can recover the ledger by getting it from other peers, so the communication cost is proportional to the storage cost. The minimum communication cost is essentially
\[
\log_2 \eta + \log_2 q. \tag{9}
\]

When there is a node failure, our proposed scheme requires \((t + 1)\) secret shares to recover the transaction block, so the recovery communication cost is dependent on the number of shares \((t + 1)\) and proportional to the storage cost:
\[
\frac{(t + 1)(\log_2 \eta + \log_2 q)}{t + 1} = \log_2 \eta + \log_2 q. \tag{10}
\]

Figure 5 shows the tradeoff between storage and recovery communication cost for a specific case. It is clearly shown that our proposed scheme can achieve lower storage cost but
5.3. Robustness to Peer Failures. In regard to the robustness to peer failures, traditional blockchain can tolerate \((n - 1)\) node failure, in the expense of large ledger stored in every node. The scheme proposed by Raman and Varshney [5] and Kim et al. [6] can only tolerate \((n/t + 1) - 1\) and \(2(n/t + 1) - 1\) failures, respectively, which is significantly smaller than the traditional blockchain. Our proposed scheme can tolerate \((n - t - 1)\) node failure, which is better compared to these two state-of-the-art schemes [5, 6]. Figure 6 shows the comparison of robustness where \(n = 900\).

Another interesting work was presented by Dai et al. [3] recently based on network coding theory. Compared to them, our proposed scheme shares the same storage cost, communication cost, and robustness. However, their scheme can guarantee the recovery of transaction block if the collected packets is no less than \(t\). It may leak information to malicious node if the collected packets is less than \(t\). On the contrary, our proposed scheme utilizes Shamir’s secret-sharing scheme, which can assure that if the collected packets is less than \(r\), no additional information is being leaked. Hence, it is more advantageous compared to the scheme proposed by Dai et al.

6. Conclusions

Traditional blockchain requires huge storage room when the number of nodes increases, which limits the scalability of this emerging technology. In this article, we proposed a low-storage scheme based on secret sharing to reduce the storage cost of blockchain. The proposed scheme is able to reduce the storage cost of traditional blockchain by \(1/t\) without introducing additional communication cost. It is also more efficient compared to the other recent studies [5, 6] that attempted to solve the same problem. Although the scheme proposed by Dai et al. [3] is as efficient as our proposed scheme, it leaks partial information and is considered less secure. The proposed scheme requires both \(r\) and \(n\) to be fixed, so it is more suitable for private and consortium blockchain systems where the number of nodes is
Robustness of different blockchain schemes.

Figure 6: Comparison of robustness (n = 900).

predetermined. This can be a limitation as the number of nodes in public blockchain is changing dynamically. In future, we plan to extend this to support public blockchain.

Data Availability

The relevant analysis data used to support the findings of this study are included in the article.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work was partly supported by the National Natural Science Foundation of China (Grant no. 11804114), the Natural Science Foundation of Fujian Province, China (Grant nos. 2017J01761 and 2018J01357), the Science and Technology project of Fujian Province (Grant no. 2019H0021), and the Science and Technology project of Xiamen Municipal (Grant no. 3502Z20173028).

References


