OPTIMIZED PRACTICAL BYZANTINE FAULT TOLERANCE (O-PBFT) ALGORITHM USING GROUPING METHOD FOR CONSORTIUM BLOCKCHAIN

JANNAH YUSOFF¹, ZARINA MOHAMAD², HASNI HASSAN³, AZNIDA HAYATI ZAKARIA@MOHAMED⁴, SHADI A ALJAWARNEH⁵

¹,²,³ Faculty of Informatics & Computing, University Sultan Zainal Abidin, Terengganu, Malaysia
⁵CIT Faculty, Jordan University of Science and Technology, Irbid, Jordan
E-mail: ¹jannahyusoff03@gmail.com, ²zarina@unisza.edu.my, ³hasni@unisza.edu.my, ⁴aznida@unisza.edu.my, ⁵saaljawarneh@just.edu.jo

ABSTRACT

Blockchain is a distributed ledger that records every transaction that has ever occurred in a system. A consensus algorithm is required in blockchain to ensure that all transactions function properly. At present, the consensus algorithm that is commonly used in consortium blockchain is Practical Byzantine Fault Tolerance (PBFT) algorithm. PBFT requires all nodes in the network to participate in the consensus process. However, PBFT still has some problems such as communication overhead, low throughput, and high latency due to increasing the number of nodes in the network. To overcome these disadvantages, an optimized PBFT (O-PBFT) algorithm is proposed. The O-PBFT algorithm used grouping method in a mid-stage (prepare stage) to reduce communication complexity and assign random Byzantine nodes to improve consensus efficiency. The consistency protocol in O-PBFT were modify from the original PBFT so that O-PBFT can reach consensus with less communication in a stable network. Experimental results show that O-PBFT algorithm reduces number of communication times between nodes, increases transaction throughput, and improves consensus efficiency compared to the original PBFT algorithm. Experimental results prove that O-PBFT algorithm can be used when many nodes are involved.

Keywords: Consensus Algorithm, Practical Byzantine Fault Tolerance (PBFT), Blockchain, Consortium Blockchain, Grouping Method.

1. INTRODUCTION

A blockchain is a decentralized, tamper-resistance database that is shared and accessible to everybody, controlled by no single entity [1]. The bitcoin system’s core technology, blockchain, was initially made public in Satoshi Nakamoto’s 2008 paper “Bitcoin: A peer-to-peer electronic cash system”. L. M Bash et al. [2] said that, according to the original Bitcoin whitepaper, this new technology was developed with the intention of enabling the development of a “peer-to-peer version of electronic cash [which] would allow online payments to be sent directly from one party to another without going through a financial institution”. Blockchain can be viewed as a novel decentralized architecture [3] and a distributed computer paradigm that stores data in encrypted chained blocks. It verifies the data using distributed consensus algorithms that ensure the security and privacy of data access. Data transmission is secured with cryptography, manipulated using self-executed program scripts (also known as smart contracts).

The structure of the blockchain is illustrated in Figure 1. As new sets of “blocks” are added to the ledger, it continues to expand. Each block contains transaction information, a timestamp, and a link to the previous block, forming a continuous chain. By hashing the previous block and linking the hash value into the current block, the continuous chain is protected. This enables a trust chain of block or tamper resistant property since the genesis block. The ledger is not managed by a single identity. Instead, each user on the network receives a copy of the entire ledger. Old blocks are maintained indefinitely, and new blocks are added to the ledger irreversibly, making it nearly impossible to tamper data by fabricating transactions, and other data.
Blockchain can be divided into three categories that are public blockchain, private blockchain, and consortium blockchain. In a public blockchain or decentralized permissionless blockchain, anyone can participate in the consensus process and contribute to receive the rewards by following the rules, and also can freely join the network [4]. An example of a public blockchain is Bitcoin. Next, private blockchain can be viewed as a centralized blockchain since a single authority or organization controls and decides who can join the consensus process. Ethereum is an example of a private blockchain. Finally, consortium blockchain combines the characteristics of both public and private blockchain, that are the low trust of public blockchain and the single highly trustable entity mode of private blockchain [5]. It is suitable for semi-closed networks that are built by different enterprises such as supply chain industries.

Performance of a blockchain system is defined by the consensus algorithms used [6]. Consensus algorithm is the key factor in maintaining the safety [7] and efficiency of blockchain. Consensus algorithms [8] originated from the famous Byzantine generals’ problems, which Lamport initially introduced in his paper “The Byzantine generals’ problems” in 1982. The Byzantine generals’ problems are described in [9]. Byzantine is the capital of the ancient eastern Roman Empire. Throughout Byzantine, there are several fiefs, each guarded by a general and his troops to defend against foreign enemies. Each general has two options when encountering enemies: attack or retreat. They can minimize casualties and win a war only when all honest generals agree on an order to attack or retreat. However, Byzantine is so vast that these generals cannot discuss the order together because they must guard their fiefs separately. Hence, signalmen are used to deliver the generals’ commands. The generals eventually make their final decision (attack or retreat) by sending their orders to the other generals and receiving orders back from them. In this scenario, we assume that the signalmen are honest. However, some of these generals are traitors, who may send wrong orders or send different orders to different generals, undermining the overall decision of the honest generals. In conclusion, the Byzantine generals’ problem can be defined as a problem of getting honest generals to reach a consensus in the presence of several traitors. Adopting this concept, consensus algorithm is used to solve the issue of maintaining data consistency in distributed systems with the presence of several failure nodes. These are some consensus algorithms introduced in the early days of blockchain systems such as Proof of Work (PoW), Proof of Stake (PoS), Delegated Proof of Stake (DPoS), and Practical Byzantine Fault Tolerance (PBFT).
Despite the fact that there are many consensus algorithms introduced and used in the blockchain, the existing consensus algorithms are mostly concerned with the public blockchain while the consortium blockchain is given the least attention [32]. This can be supported by [33] and [34] that said, at present, most of the researches on consensus algorithms are focused on the public blockchain and there are few consensus algorithms that have been developed for consortium blockchain.

Consortium blockchain is a permissioned blockchain, in which the primary nodes are pre-specified by the participants which are composed of many parties. Thus, whoever wants to access to the ledger needs to be a member of any organization. PBFT algorithm is widely used in consortium blockchain. However, PBFT has poor and low scalability because of the communication complexity [9]. The communication complexity issue in the PBFT algorithm is due to increased communication with a greater number of nodes. Therefore, this paper proposes an optimized PBFT (O-PBFT) that improved from the original PBFT algorithm. To overcome the original PBFT algorithm disadvantages, we are using the grouping method (the nodes will achieve consensus within their group only) in the prepare stage that can improve scalability and performance while ensuring fault tolerance.

The rest of the paper is organized as follows. Section 2 presents an overview of the PBFT algorithm, its disadvantages, and improvement methods in PBFT by other research papers. Section 3 introduces our O-PBFT consensus algorithm while performance analysis, and experimental results are presented, and discussed in section 4. Finally, in section 5; we provide the conclusions together with recommendations for future work.

2. PRACTICAL BYZANTINE FAULT TOLERANCE (PBFT) ALGORITHM

2.1 Overview of the PBFT Algorithm

Practical Byzantine Fault Tolerance (PBFT) algorithm was proposed by Castro and Liskov to solve the General Byzantine problem in a paper that was published in 1999. The PBFT’s main goal is to replicate Byzantine state machines in a useful way such that the Byzantine fault can be tolerated [10] and the algorithm can be used in real-world system applications [11]. The process of PBFT algorithm is divided into three phases that are pre-prepare, prepare, and commit [12]. In each phase, a node moves on to the following phase if it receives votes from more than two-thirds of all nodes [8]. PBFT algorithm provides a fault tolerance of (n—1)/3 under the premise of ensuring activity and safety. The maximum number of malicious nodes, f must not exceed 1/3 of the total number of nodes, n so that

\[ n > 3f + 1 \]  

PBFT algorithm flow is illustrated in Figure 2 [13], where Node 3 is considered to be the Byzantine node, while Nodes 0 (primary node), 1, and 2 are the normal nodes. The three phases in PBFT are described as follows:

1) Request: The client sends a request to the primary node.
2) Pre-prepare: The primary node checks, and processes the client’s request and broadcasts a pre-prepare message to the nodes. Each node receives and verifies the validity of the pre-prepare message. Once it is authenticated, a node will accept the request and start the prepare phase.
3) Prepare: The nodes broadcast prepared messages to each other and also received prepared messages from others and checked its validity. When a node gets 2f valid prepared messages from different nodes, the prepare phase is finished, where f is the number of Byzantine nodes in the system.
4) Commit: Each node broadcasts a commit message to others for validation. Once the number of the commit message including itself that is received by a node in the prepare phase is greater than 2f+1, it will send a reply message to the client.
5) Reply: When the client has received f+1 of the same reply message, the consensus is achieved.

2.2 Disadvantages of the PBFT Algorithm

Generally, PBFT algorithm is widely used in a consortium blockchain because it significantly improves the performance of the blockchain consensus. This algorithm has a better overall performance and is more suitable for most customized application scenarios due to its advantages of low time delay, low energy consumption, no bifurcation, and resolution of Byzantine fault tolerance. However, according to [9, 14-15], PBFT has poor and low scalability. To reach a consensus in the PBFT algorithm, all nodes must communicate with one another. This results in increased communication with a greater number of
nodes, resulting in increased communication overhead. Because of this, PBFT does not have good scalability. This is also supported by M. Vukolic [16] iterating that the PBFT algorithm processes transaction requests at high speed, but the overhead resulting from communications limits their scalability. Wei Liu et al. [17] concurred that PBFT has the advantages of high efficiency, and fast feedback as a consensus algorithm in consortium blockchain. However, when the distributed system is scaled up, PBFT is inefficient due to its high resource consumption and latency. As the number of nodes increases in the PBFT system, the communication costs required to reach a consensus in the mesh network increase exponentially, resulting in a scalability problem. The authors of [18] said that the blockchain can only scale to a few tens of nodes due to the high communication complexity. When the number of nodes in the network exceeds this threshold, the transaction confirmation delay of the PBFT algorithm increases significantly, and the throughput will be greatly reduced.

2.3 Improvement Method in the PBFT Algorithm by Other Research Papers

The PBFT algorithm has undergone extensive study and improvements by many scholars [19] along with the development of today’s consortium blockchain. The three-phases consensus process between the nodes of the PBFT algorithm produces most of the communication overhead. Currently, the main improvement ideas are as follows. X. F. Liu [20] proposed a simplified PBFT three-stage protocol that simplifies the three-phase consensus of PBFT to two-phase consensus, which reduces the communication overhead, and improves the consensus efficiency. However, reducing the consensus phase will compromise the security and consistency of the algorithm. In another method, a reputation model is used to assess the trust of nodes based on their behavior, and nodes with high trust of nodes are elected as primary nodes or nodes with high trust value of nodes are chosen to form a consensus group as opposed to all nodes participating in consensus. To improve Byzantine fault tolerance, [21] proposed an optimized PBFT based on the feature trust model. In [22], a consensus group is created from nodes with a high trust degree. Still, the credibility model requires multiple iterations of credibility assessment and at the end of each consensus, the nodes must be assessed for trust. This aggregates the node voting results using a digital signature approach before sending the statistical data to each node individually. SBFT proposed in [23] assigns collectors in the blockchain network to gather and relay each node’s votes throughout the three-phase consensus to minimize direct communication between nodes. Based on this technique, direct contact between nodes can be avoided but as the number of nodes in the network increases, the aggregated signature grows too large, necessitating high network performance. In another method, a grouping approach is used, based on node attributes (node hardware, performance, network communication capabilities, geographic location, etc). The nodes are divided into multiple groups by clustering algorithms, with intragroup consensus coming first and subsequently intergroup consensus. To perform intragroup and intergroup consensus by clustering and hierarchical dividing the properties of large-scale network nodes, an improved PBFT consensus mechanism based on K-medoids is proposed in [24]. In [25], a network self-clustering-based approach is proposed to group nodes based on their communication capabilities, and nodes are grouped based on their geographical locations in [26]. Based on these papers, we can conclude that grouping approach can reduce communication overhead and improve consensus efficiency, hence throughput of a system will be significantly improved.

3. OPTIMIZED PRACTICAL BYZANTINE FAULT TOLERANCE (O-PBFT) ALGORITHM

3.1 Parameters

The main parameters involved in optimized PBFT (O-PBFT) consensus algorithm in this paper are shown in Table 1.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>total nodes</td>
</tr>
<tr>
<td>f</td>
<td>total faulty nodes (expected)</td>
</tr>
<tr>
<td>d</td>
<td>total groups</td>
</tr>
<tr>
<td>k</td>
<td>nodes in groups</td>
</tr>
<tr>
<td>group_prepare</td>
<td>total nodes for the first layer (prepare phase)</td>
</tr>
<tr>
<td>total_faulty_nodes_first_layer</td>
<td>total faulty nodes for the first layer (prepare phase)</td>
</tr>
<tr>
<td>group_commit</td>
<td>total nodes for the second layer (commit phase)</td>
</tr>
<tr>
<td>total_faulty_nodes_second_layer</td>
<td>total faulty nodes for the second</td>
</tr>
</tbody>
</table>
3.2 Flow of O-PBFT Algorithm

The PBFT consensus algorithm is a state machine replication algorithm [27]. The proposed algorithm is designed to use less resources to achieve the same outcome as the original PBFT algorithm. According to the PBFT algorithm [28], the algorithm model is modified based on the organization and properties of the nodes. First, we assume that the entire blockchain network is available where each node can send any messages to any other nodes in the blockchain network in real time. Next, the assumptions are there are an unknown number of Byzantine nodes in the network and the entire network system works in a permissioned environment, requiring identity verification of the node before it can participate in the consensus protocol. The Byzantine nodes, \( f \) is at most

\[
f = \left(\frac{n - 1}{3}\right)
\]  

(2)

O-PBFT algorithm flow is illustrated in Figure 3 [29] with the following processes:

1) Request: The client sends a request to the primary node.
2) Pre-prepare: The primary node checks, and processes the client’s request, and then broadcasts a pre-prepare message to the nodes. Each node receives and verifies the validity of the pre-prepare message. Once it is authenticated, a node will accept the request, and start the prepare phase.
3) Prepare: In this phase, the node does not need to communicate with all other nodes, but within their groups only. This happens if the total number of nodes in the blockchain, \( n \) is more than or equal to \( x \), and \( x \) is the preset number of nodes. The primary node will send a client’s request to \( d \) representative nodes where \( d \) represents the number of groups in the network. If the total number of nodes in the blockchain is less than to the number of \( x \), there is no need to form groups. In this case, PBFT can be directly used in the blockchain. When each node receives \( 2f \) valid prepared messages consistent with itself and get acknowledgement within their groups, it means that all nodes in the groups have reached consensus, and they will move to the commit phase.
4) Commit: The group (node) broadcasts a commit message to others for validation. Once the number of the commit messages including itself received by a group (node) in the prepare phase is greater than \( 2f+1 \), it will send a reply message to the client. The decision is made based on majority groups which are 2/3 from all groups. The concept to achieve the majority node in this proposed algorithm (two-third) is still the same with the original algorithm, PBFT.
5) Reply: When the client has received \( f+1 \) of the same reply message, the consensus is achieved.

3.3 Grouping in Prepare and Commit Stage

The communication complexity problems in the original PBFT algorithm are due to the large communication among the nodes to achieve a consensus. To reduce the communication cost, intuitively, one can construct an optimized PBFT by refraining the communication within their groups only. A sub-consensus can be reached for each group, and the consensus is reached when more groups reach their sub-consensus. Based on Figure 4 [30], grouping method starts in the prepare stage. The nodes will be grouped into a few groups, sequentially where the first node will be grouped in group 1, followed by group 2, group 3, and so on. These nodes in the groups will communicate and achieve the consensus within their groups. In the commit stage, the group (node) will broadcast a commit message to another group (node) for validation, get the decision between them. When the decision is made based on majority that are 2/3 from all groups, the consensus is achieved.

4. EXPERIMENT

A simulation platform was proposed to analyze the performance of O-PBFT. The code is written in the Python language (version 3.9.10). This code is improved from [31]. For the accuracy of the experiment, the network environment is excluded and all the consensus nodes are running on the same host. The software and hardware configuration are shown in Table 2.
4.1 Latency

Latency is one of the important indexes of consensus protocols. Latency represents the time difference from transaction submission to transaction confirmation. Figure 5 shows the relationship between the number of nodes and the consensus latency. The latency of the PBFT algorithm increases rapidly as the number of nodes grows. In contrast, the variation of latency based on the proposed algorithm is more stable as the number of nodes increases.

![Figure 5: Latency PBFT and O-PBFT](image1)

4.2 Throughput

Throughput is also another important indexes for measuring consensus algorithms. Throughput represents the number of transactions that the network can process per second, which is, the number of transactions generated over time.

![Figure 6: Throughput PBFT and O-PBFT](image2)
during the test. Throughput performance is usually expressed as transactions per second (TPS) when presenting performance test results. Figure 6 shows the comparison of throughput of PBFT and O-PBFT with the same number of nodes. It can be observed that, as the number of nodes in the network increases, the throughput of PBFT and O-PBFT will decrease. However, the throughput of the O-PBFT algorithm is much higher than the original PBFT algorithm.

4.3 Communication Complexity

Communication complexity is an important index of the consensus protocols [15]. It can be reflected by the number of communications times. Table 3 presents a comparison of communication complexity of each phase (pre-prepare, prepare, commit) of the original PBFT algorithm and the optimized PBFT algorithm, O-PBFT. To show the differences more clearly between them, the communication complexity is compared at different numbers of nodes. For example, there are 80 nodes in the network, where \( n = 80 \). If each group has 4 nodes, there are 26 groups; \( d = 26 \), and \( k = 4 \). According to Table 3, the complexity of PBFT is 12719 while the complexity of O-PBFT is only 985.

Table 3: Comparison of Communication Times PBFT & O-PBFT

<table>
<thead>
<tr>
<th>Phases</th>
<th>PBFT</th>
<th>O-PBFT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-prepare</td>
<td>( n - 1 )</td>
<td>( n - 1 )</td>
</tr>
<tr>
<td>Prepare</td>
<td>( n \times (n - 1) )</td>
<td>((k - 1)^2 \times (n / 3))</td>
</tr>
<tr>
<td>Commit</td>
<td>( n \times (n - 1) )</td>
<td>((d - 1) \times (n / 3))</td>
</tr>
<tr>
<td>Total</td>
<td>(2n^2 - n - 1)</td>
<td>((n - 1) + ((k^2n - 2kn + n) / 3) + ((d(n - n)) / 3))</td>
</tr>
</tbody>
</table>

When the number of nodes in the blockchain network varies between 10 and 100, the communication complexity and communication times of PBFT and O-PBFT are shown in Table 4 and Figure 7. It can be observed that the communication complexity of the PBFT algorithm sharply increases with the number of nodes, reflecting the algorithm is \(2n^2 - n - 1\) message complexity. Meanwhile, the communication complexity of the O-PBFT algorithm does not increase much as the number of nodes increases. In particular, when there are 100 nodes in the blockchain network, the communication complexity of the PBFT algorithm is almost 20 times more than the communication complexity of the O-PBFT algorithm. It demonstrates that our proposed algorithm significantly reduces the number of communication times between nodes compared to the original PBFT algorithm.

Table 4: Communication Times PBFT & O-PBFT

<table>
<thead>
<tr>
<th>No. of Nodes</th>
<th>PBFT</th>
<th>O-PBFT</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>189</td>
<td>45</td>
</tr>
<tr>
<td>20</td>
<td>779</td>
<td>112</td>
</tr>
<tr>
<td>30</td>
<td>1769</td>
<td>209</td>
</tr>
<tr>
<td>40</td>
<td>3159</td>
<td>319</td>
</tr>
<tr>
<td>50</td>
<td>4949</td>
<td>449</td>
</tr>
<tr>
<td>60</td>
<td>7139</td>
<td>619</td>
</tr>
<tr>
<td>70</td>
<td>9729</td>
<td>792</td>
</tr>
<tr>
<td>80</td>
<td>12719</td>
<td>985</td>
</tr>
<tr>
<td>90</td>
<td>16109</td>
<td>1229</td>
</tr>
<tr>
<td>100</td>
<td>19899</td>
<td>1465</td>
</tr>
</tbody>
</table>

Figure 7: Communication PBFT and O-PBFT
5. CONCLUSIONS AND FUTURE WORK

In this paper, an optimized PBFT (O-PBFT) algorithm is proposed to optimize the problem of the original Practical Byzantine Fault Tolerance (PBFT) algorithm when greater nodes are involved. O-PBFT used a grouping method in the prepare stage to reduce communication times between nodes and assign random byzantine nodes to regulate the process fairly. This paper demonstrates that the proposed O-PBFT shows better performance in terms of latency, throughput and communication complexity compared with the original PBFT algorithm. However, the cons of this proposed algorithm are still available but we still can try to improve it because the consensus algorithm that is specially designed for each scenario is still very rare. For future work, we will focus on developing a more efficient and effective optimization model that can decide exactly on the number of nodes needed in a group to optimize the consensus.

ACKNOWLEDGEMENT

This work is supported by Matching Grant (UniSZA/2023/JORDAN (JUST)/001) Universiti Sultan Zainal Abidin (UniSZA), Malaysia and Jordan University of Science and Technology (JUST), Jordan.

REFERENCES:


[10] Alsunaidi SJ, Alhaidari FA. “A survey of consensus algorithms for blockchain technology”. In 2019 International Conference on Computer and Information Sciences (ICCIS) 2019 Apr 3 (pp. 1-6). IEEE.


Figure 2: PBFT Algorithm Flow [13]

Figure 3: O-PBFT Algorithm Flow [29]
Figure 4: O-PBFT Algorithm Grouping [30]