Efficient User Inspection Algorithm Based on Dual Bloom Filters Oriented for Blockchain Data Management Systems

Wenlong Yi
School of Software Engineering
Jiangxi Agricultural University
Nanchang, China

Qiude Li
School of Software Engineering
Jiangxi Agricultural University
Nanchang, China

Hua Yin
School of Software Engineering
Jiangxi Agricultural University
Nanchang, China

Hao Tang
School of Software Engineering
Jiangxi Agricultural University
Nanchang, China

Yingding Zhao*
School of Software Engineering
Jiangxi Agricultural University
Nanchang, China
*Corresponding author:
zhaoyingding@163.com

***Abstract*—With the increasing application of blockchain technology in decentralized data management systems (DMS), the prevailing problems of poor query performance relative to its conventionally centralized counterparts are in urgent need of reasonable solutions. Bloom filters are efficient data query frameworks that adopt multiple hash functions to map the target database into one-dimensional arrays (one bit per array cell), leading to high-efficiency information extraction and fulfilling the urgent optimization needs of the blockchain's query performance. In light of the above background, dual Bloom filters were proposed herein to accelerate the inspection speed of user information in a blockchain DMS. Taking into account the different scenarios for user registration and user login, a comparative experiment was conducted between the conventional methods and the proposed algorithm, with the latter outperforming the former, as revealed by the results. The proposed algorithm has proven to be capable of swiftly completing not only the duplicate detection of usernames for newly-registered accounts but also the parallel query for the username and password during user login. This functionality has greatly reduced noise from irrelevant information and improved the query performance of blockchain-based hyperledgers.**

Keywords—blockchain, bloom filter, hyperledger, query efficiency, data management

I. INTRODUCTION

Since data are subject to tampering and damaging in the traditional information management system, data reliability can hardly be proven. Blockchain is a decentralized and distributed hyperledger based on the technological development of Bitcoin. Data stored in the blockchain cannot be tampered while data integrity can be also tracked for the Hash address value of the previous block is stored in each block in the data structure of the blockchain, except for the genesis block [1]. As the size of data access in the blockchain is enlarged gradually, it is not as strong as the traditional centralized database management system for data processing delay, functional scalability[2][3]. In particular, it is essential to improve the performance of blockchain-based DMS, integrating with other real-time acquisition systems [4][5]. User login information management of the commonly-used block in the DMS is taken as an example for specific explanation. When a user logs in to the system, all user information on the blockchain should be traversed using the identity verification algorithm, leading to massive expense with the decline in the user's experience of using the system. In this study, a dual Bloom filter algorithm is proposed to reduce the time cost of user inspection of the DMS through eliminating the traversal of invalid user data. It is worth noting that Bloom filter is a probabilistic data structure that has been applied to the query optimization of blockchain systems [6]–[9].

This study was supported by the National Key Research and Development Program of China (Grant No. 2020YFD1100603) and the National Natural Science Foundation of China (Grant No. 61762048).

II. METHOD

A. Concepts

Bloom filter is to map user information to be queried using multiple hash functions that are not interdependent into a single array of bits. As can be seen in Fig. 1, when the mapped value is incompletely corresponding to the array element value 1, the information does not exist. However, a false rate (that is, false positives) might be found, or a certain deviation might exist in judging the existence of user information due to possible hash collisions.

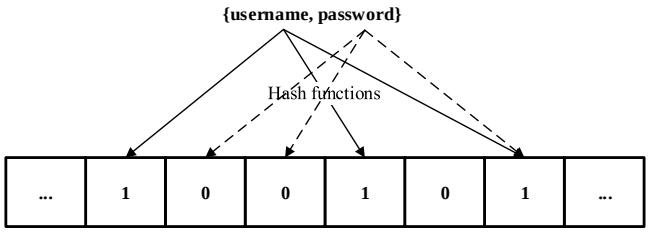


Fig. 1. Concept map of the Bloom filter

B. Algorithm design based on dual Bloom filters

User inspection algorithm in the blockchain DMS is principally applied to two scenarios, user registration and login. Precisely, a bloom filter is adopted to test whether the username chosen by the user is taken during registration, as shown in Fig. 2. If it is not, then the new user can take the

username as the login account. In the algorithm, username and password are mapped to the Bloom filter using hash functions. And successful registration is displayed in the system after the account information is stored in the blockchain. Otherwise, the new user shall take another username for re-registration.

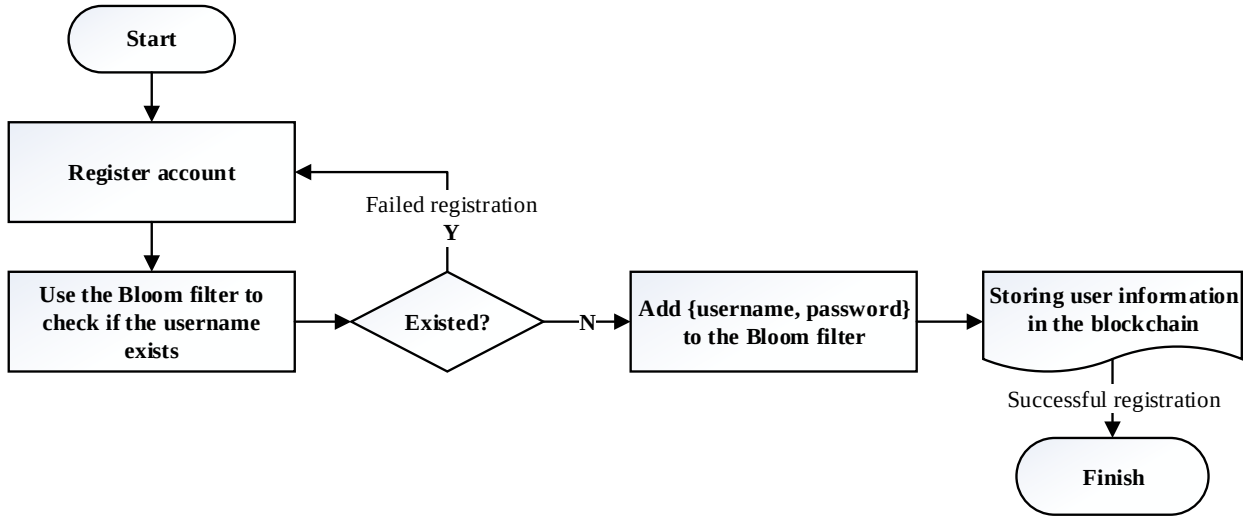


Fig. 2. User registration algorithm

In the scenario of user login, the user firstly enters a username and password, as shown in Fig. 3. Next, the first username detection is performed in the algorithm bloom filter. If the username exists, the second password detection will be conducted before comparing the username and password with the data stored in the blockchain. Then, if it exists, successful login will be displayed. Otherwise, the user is prompted to re-enter the username and password.

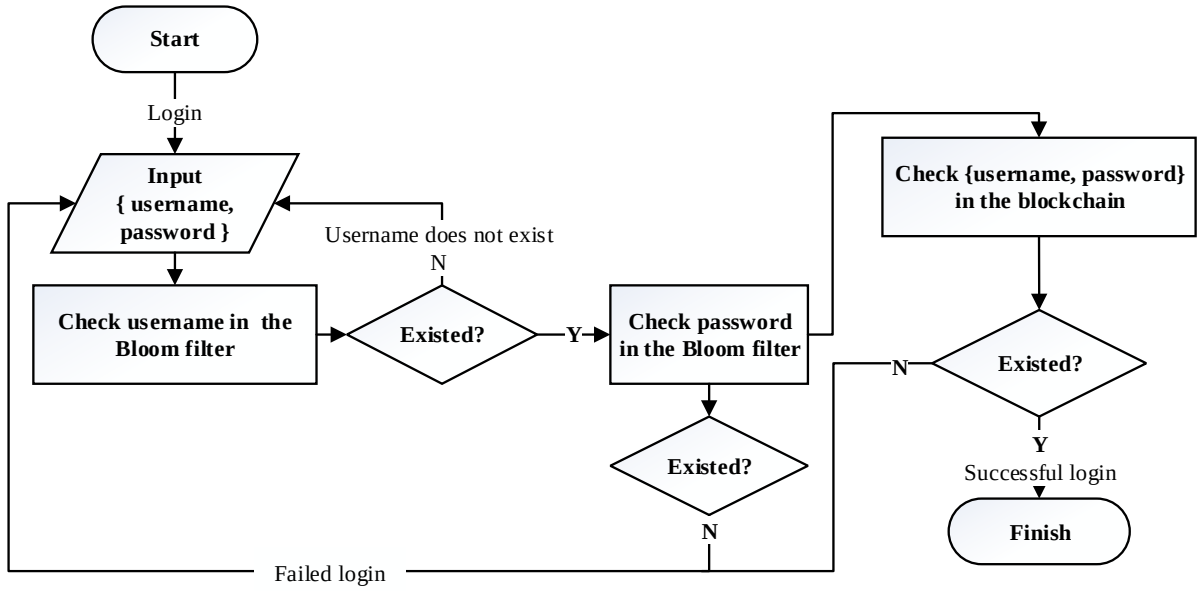


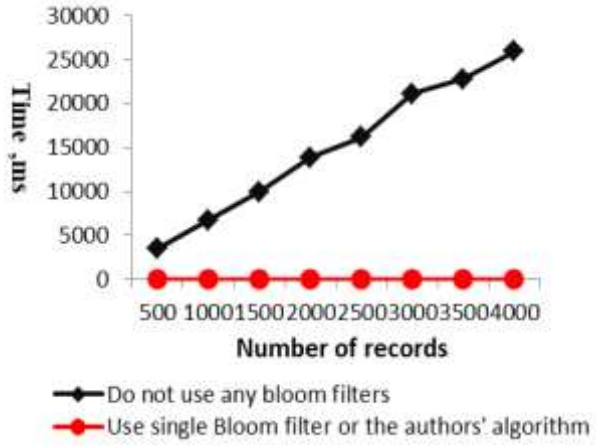
Fig. 3. User login algorithm

III. RESULTS AND DISCUSSION

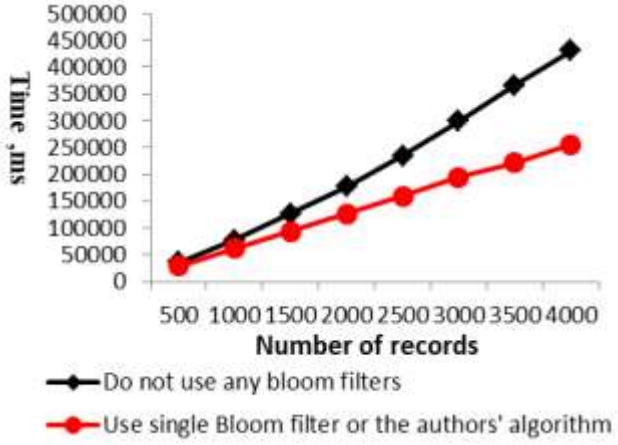
To verify the effectiveness of the algorithm in user registration and login, a cloud server equipped with CPU: 1 core, 2GB memory, Ubuntu/16.04 operating system in the development environment of V2.2 Fabric blockchain framework, Golang programming language, and V18.09.7 Docker container was selected. To begin with, the Bloom filter algorithm was implemented using go language. Secondly, smart contracts were deployed by efficiently setting up Fabric basic network and Fabric-SDK-go using Docker. At last, a 2-organization and 4-node cluster network was built. In addition, Murmurhash3 that is an efficient non-cryptographic hash function that has been extensively proven was adopted since frequent hash operations should be performed in the Bloom filter on the data to be queried. At the same time, a multitude of independent functions that can be generated through initializing seeds can fulfill the needs of Bloom filters with Murmurhash3 function as the hash function [10]. In the simulation test, six Murmurhash3 hash functions were generated. Also, a one-dimensional array length of 64MB was constructed, and the probability of erroneous judgment was set to 0.01. Based on this, changes in the response efficiency of the blockchain data

management system when users of 500 to 8000 data records log in were simulated.

When a user registers an account, time cost will be consistent in using a single or a dual bloom filter as testing the existence of user name is required in the system. As can be observed from Fig. 4, if the account existed, it can be verified by the bloom filter efficiently, presenting a remarkable enhancement in performance. If the account does not exist, using the bloom filter to inspection also witnesses a certain advantage.



(a)



(b)

Fig. 4. User account registration. (a) The username existed; (b) the username does not exist

The user might experience the username does not exist, the username is correct with the wrong password, and the correct username and password during login. As shown in Fig. 5–7, The results of “not using any bloom filter”, “using a single bloom filter” and "using dual bloom filters" were compared during the test. If the username does not exist, the performances of dual bloom filters and a single bloom filter are similar and far superior to that of not using any bloom filters. If the username is correct with the wrong password, the performance of dual bloom filters is much better than that of the other two cases. And the performances of the three cases are similar in the case of correct username and password.

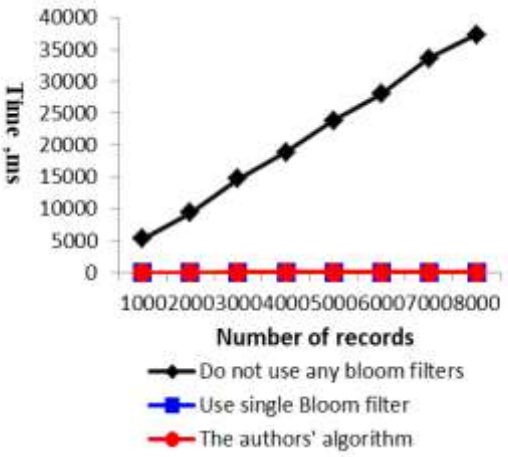


Fig. 5. The username does not exist

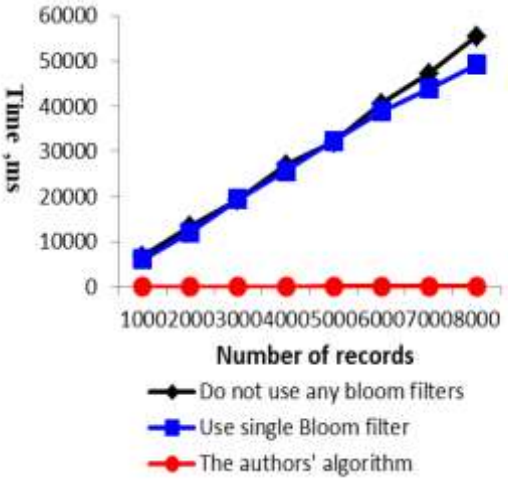


Fig. 6. Wrong password for the username

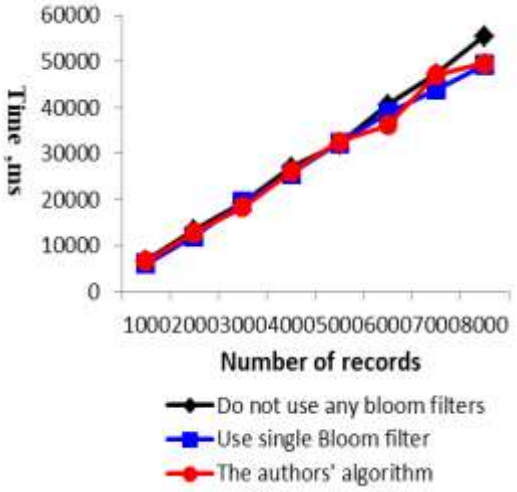


Fig. 7. Correct username and password

In this work, an algorithm based on dual bloom filters was proposed against user formation inspection in the blockchain data management system, which can decrease the number of direct data traversal in the system through adding a filtration of password and matching with data in the blockchain upon doubling operations of eliminating invalid information. Also, data security can also be ensured since the

Bloom filter maps merely the hash function value, rather than the data information, of account and password to be queried into a one-dimensional data. Nevertheless, the same data structure applied to username and password hash mapping operations in the proposed algorithm using dual bloom filters undoubtedly increases the probability of hash collisions. In that case, two data structures or increased length of the one-dimensional array can be established in the follow-up work based on the practical demands, so as to reduce the probability of erroneous judgment.

IV. CONCLUSION

Since Bloom filter is a probabilistic data structure, which is completely reliable for judging that an element does not exist. However, erroneous judgment might be generated in terms of judging an account because of Hash collisions. Therefore, the proposed algorithm is first to eliminate invalid account before judging the validities of the account and password. Based on this, account information is compared in the blockchain. Account validation performing upon the elimination of all error information can contribute to reducing the interference of irrelevant information, and improving the query efficiency of the Hyperledger based on the blockchain technology on the premise of safeguarding the efficiency and accuracy of information inspection.

REFERENCES

[1] Junhui W., Tuolei W., Yusheng W., Jie C., Kaiyan L. and Huiping S. Improved Blockchain Commodity Traceability System Using Distributed Hash Table. 2020 Chinese Automation Congress (CAC), Shanghai, China, 2020, pp. 1419-1424, doi: 10.1109/CAC51589.2020.9326639.

[2] Hao J., Sun Y. and Luo H. A safe and efficient storage scheme based on blockchain and IPFS for agricultural products tracking. Journal of Computers, 2018, vol. 29. No. 6, pp. 158-167.

[3] Raikwar M., Gligoroski D. and Velinov G. Trends in Development of Databases and Blockchain. 2020 Seventh International Conference on Software Defined Systems (SDS), Paris, France, 2020, pp. 177-182, doi: 10.1109/SDS49854.2020.9143893.

[4] Nikouei S. Y., Xu R., Nagothu D., Chen Y., Aved A. and Blasch E. Real-Time Index Authentication for Event-Oriented Surveillance Video Query using Blockchain. 2018 IEEE International Smart Cities Conference (ISC2), Kansas City, MO, USA, 2018, pp. 1-8, doi: 10.1109/ISC2.2018.8656668.

[5] Huang Sihan et al. Blockchain-based data management for digital twin of product. Journal of Manufacturing Systems, 2020, vol. 54, pp. 361-371.

[6] Luo L, Guo D, Ma R T B. et al. Optimizing Bloom Filter: Challenges, Solutions, and Comparisons. IEEE Communications Surveys & Tutorials, 2019, vol. 21, iss. 2, pp. 1912-1949.

[7] Singh Amritpal, et al. Bloom filter based optimization scheme for massive data handling in IoT environment. Future Generation Computer Systems, 2018, vol. 82, pp. 440-449.

[8] Christen P., Ranbaduge T., Vatsalan D. and Schnell R. Precise and Fast Cryptanalysis for Bloom Filter Based Privacy-Preserving Record Linkage. IEEE Transactions on Knowledge and Data Engineering, 2019, vol. 31, no. 11, pp. 2164-2177, doi: 10.1109/TKDE.2018.2874004.

[9] Lee, Jungwon, Hayoung Byun, and Hyesook Lim. Dual-load Bloom filter: Application for name lookup. Computer Communications, 2020, vol. 151, pp. 1-9.

[10] Murmur Hash, [online] URL: https://sites.google.com/site/murmurhash/.