Fourteenth International Workshop on
Parallel Programming Models and Systems Software for
High-End Computing (P2S2), 2021

FMSM: A Fuzzy Multi-keyword Search Scheme for Encrypted Cloud Data based on Multi-chain Network

Heng He, Chengyu Liu, Xiaohu Zhou, Ke Feng

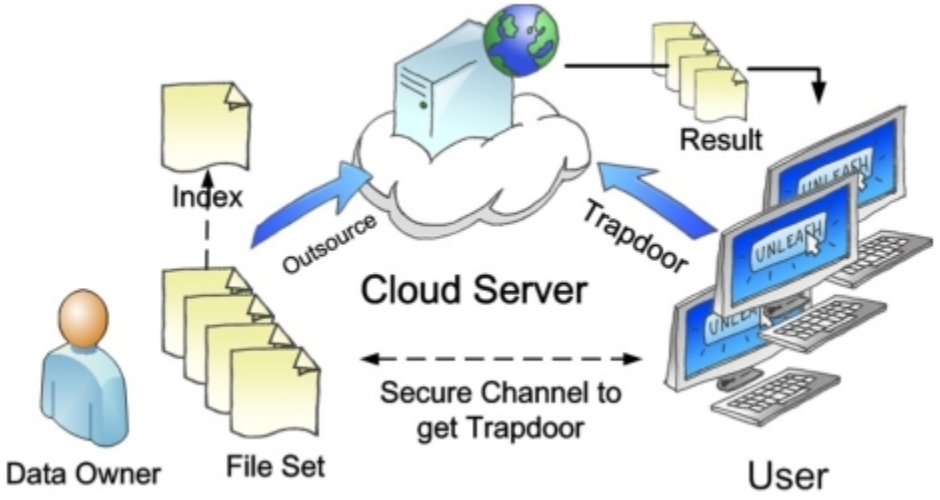
School of Computer Science and Technology, Wuhan University of Science and Technology, Wuhan, China.
Hubei Province Key Laboratory of Intelligent Information Processing and Real-time Industrial System, Wuhan, China.
School of Computing, Engineering and Built Environment, Birmingham City University, Birmingham, UK.

Agenda

- **Background and Motivation**
- **Our Method**
- **Experiments**
- **Conclusion**

• Background and Motivation

Traditional searchable encryption schemes



The main steps:

Step 1: DO encrypts plaintext file sets into ciphertext file sets with a symmetric encryption algorithm, generates secure indexes, and sends them to the cloud server.

Step 2: DU generates a trapdoor according to search keywords and sends it to the cloud server.

Step 3: The cloud server matches the trapdoor uploaded by DU and secure indexes, and sends the retrieved files to DU.

Step 4: DU uses symmetric keys to decrypt ciphertext files.

The main concerns:
Cloud server security

Is the cloud server absolutely secure and trusted?

Most existing searchable encryption schemes often assume that the cloud server is “honest but curious”, that it will execute the programs correctly, and never actively try to deviate from the pre-defined protocol. In fact, the cloud servers may return incomplete or mismatched search results for many reasons (e.g., expose to accidental attacks and malicious control), and centralized cloud servers can perform trapdoor tests and ciphertext matching without restriction and are too powerful to be monitored.

Although a few works have been desinged to adress the above problems, these techniques still exist some limitations, such as detecting cheating behaviors or inefficient.

In summary, cloud servers cannot be considered as trusted third parties.

Blockchain

The blockchain is a bookkeeping technology that is jointly maintained by multiple parties. It empowers data secure transmission, access, and consistent storage. Every block in the blockchain is a permanent record book of transactions and the consensus mechanism ensures the consistency of the entire blockchain network. So, it has features such as decentralization, verifiability and immutability.

Based on the above features, blockchain can be used as a trusted third party to perform search work instead of cloud servers.

Related works are [20]-[25].

[20] Peng Jiang, Fuchuan Guo, Kaitai Liang, Jianchang Lai, and Qiaoyan Wen. 2020. Searchain: Blockchain-based private keyword search in decentralized storage. Future Generation Computer Systems (FGCS)107 (2020), 781-792.

[21] Suhui Liu, Jiguo Yu, Yinhao Xiao, Zhiguo Wan, Shengling Wang, and Biwei Yan. 2020. BC-SABE: Blockchain-Aided Searchable Attribute-Based Encryption for Cloud-IoT. IEEE Internet of Things Journal (IOTJ) 7, 9 (2020), 7851-7867.

[22] Shan Jiang, Jiannong Cao, Julie A. McCann, Yanni Yang, Yang Liu, Xiaoqing Wang, and Yuming Deng. 2019. Privacy-preserving and efficient multi-keyword search over encrypted data on blockchain. In IEEE International Conference on Blockchain (Blockchain), 405-410.

[23] Huige Li, Fangguo Zhang, Jiejie He, and Haibo Tian. 2017. A searchable symmetric encryption scheme using blockchain. CoRR abs/1711.01030 (2017).

[24] Shengshan Hu, Chengjun Cai, Qian Wang, Cong Wang, Xiangyang Luo, and Kui Ren. 2018. Searching an encrypted cloud, meets blockchain: A decentralized, reliable and fair realization. In IEEE Conference on Computer Communications (INFOCOM), 792-800.

[25] Biwen Chen, Libing Wu, Huaqun Wang, Lu Zhou, and Debiao He. 2020. A Blockchain-Based Searchable Public-Key Encryption with Forward and Backward Privacy for Cloud-Assisted Vehicular Social Networks. IEEE Transactions on Vehicular Technology (TVT) 69, 6 (2020), 5813-5825.

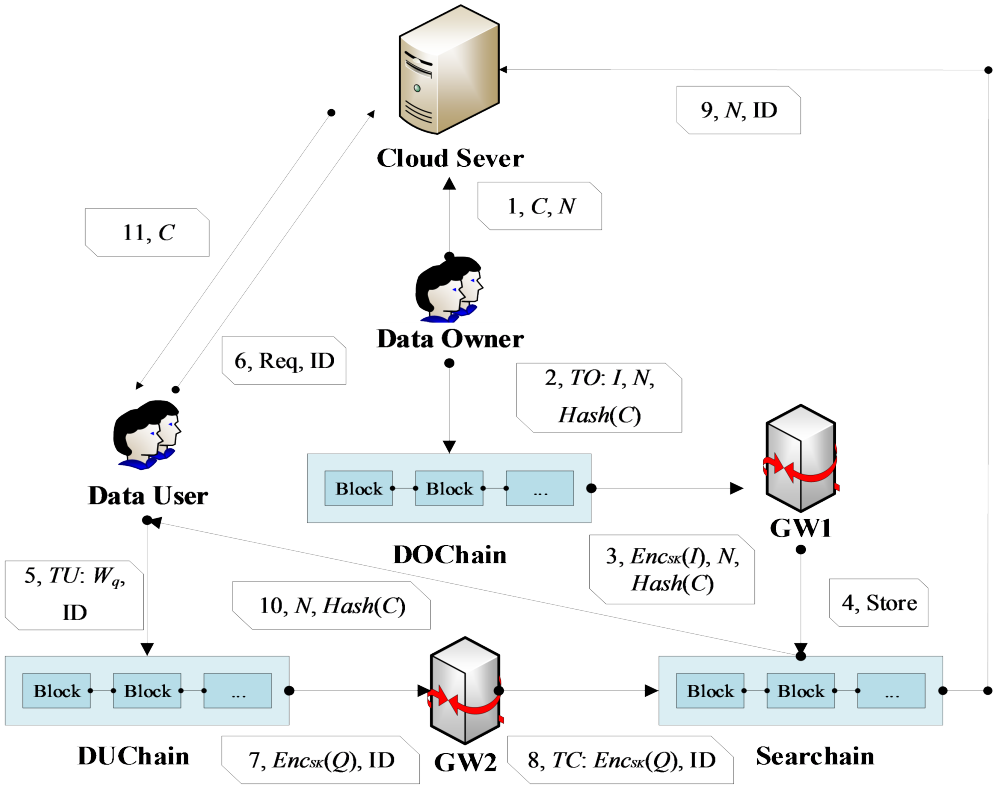
Our contributions

- Inspired by Chen et al. [25], we introduce blockchain in ciphertext search scheme to solve the trustworthiness problem of the third party and put the search task into the blockchain to ensure the correctness of the search results. We choose Hyperledger Fabric as the platform for experiments and design a multi-chain architecture. Different chains isolate participants and data (including chaincodes), which effectively protects private data and improves the efficiency of parallel processing of data and the utilization of data storage space.
- The MinHash algorithm is introduced to build the indexes on the basis of Wang's scheme. Reducing the dimensionality of the keyword set through the MinHash algorithm and calculating the similarity of the vectors based on Jaccard distance, which can greatly improve the search efficiency and reduce the space overheads without affecting the search accuracy since the keywords are processed as 01 vectors.
- Counting bloom filter [27] is used as the index structure, its counts are used to indicate keyword weights. Keywords that appear more frequently in a file have higher weights in the corresponding index structure. In the returned set of ranked results, files with higher keyword weights are ranked higher, thus more accurate results can be returned.

• Our Method

Overall structure of FMSM

- Overall structure design:



Algorithm design

***KeyGen*(*m*)**: The security key $SK = (M_1, M_2, S)$ is generated by a security parameter m, M_1, M_2 are invertible matrices of order m, $S = \{0, 1\}^m$ is an m-bit vector.

***BuildIndex*(*D, SK, H*)**: Input file D, SK and $H = \{h_j | h_j: \{0, 1\}^{26\times26} \rightarrow \{0, 1\}^m,\ j = \{1, 2, \ldots, k\}\}$, output index vector I.

***IndexEnc*(*SK, I*)**: Input I and SK, output $Enc_{SK}(I) = (\square\square_1^T I', \square\square_2^T I'')$ as the file secure index vector.

***Trapdoor*(*q, SK, H*)**: Input q, SK, H, output the query vector Q.

***QueryEnc*(*SK, Q*)**: Input SK, Q, output $Enc_{SK}(Q) = (\square\square_1^{-1} Q', \square\square_2^{-1} Q'')$ as the trapdoor.

***Search*($Enc_{SK}(I)$, $Enc_{SK}(Q)$)**:

$$(M_1^T I')^T \square M_1^{-1} Q' + (M_2^T I'')^T \square M_2^{-1} Q'' = I'^T \square Q' + I''^T \square Q''$$

If $S[t] = 1$,

$$I'[t]^T Q'[t] + I''[t]^T Q''[t] = I[t](Q[t]/2 + r') + I[t](Q[t]/2 - r') = I[t]Q[t]$$

If $S[t] = 0$,

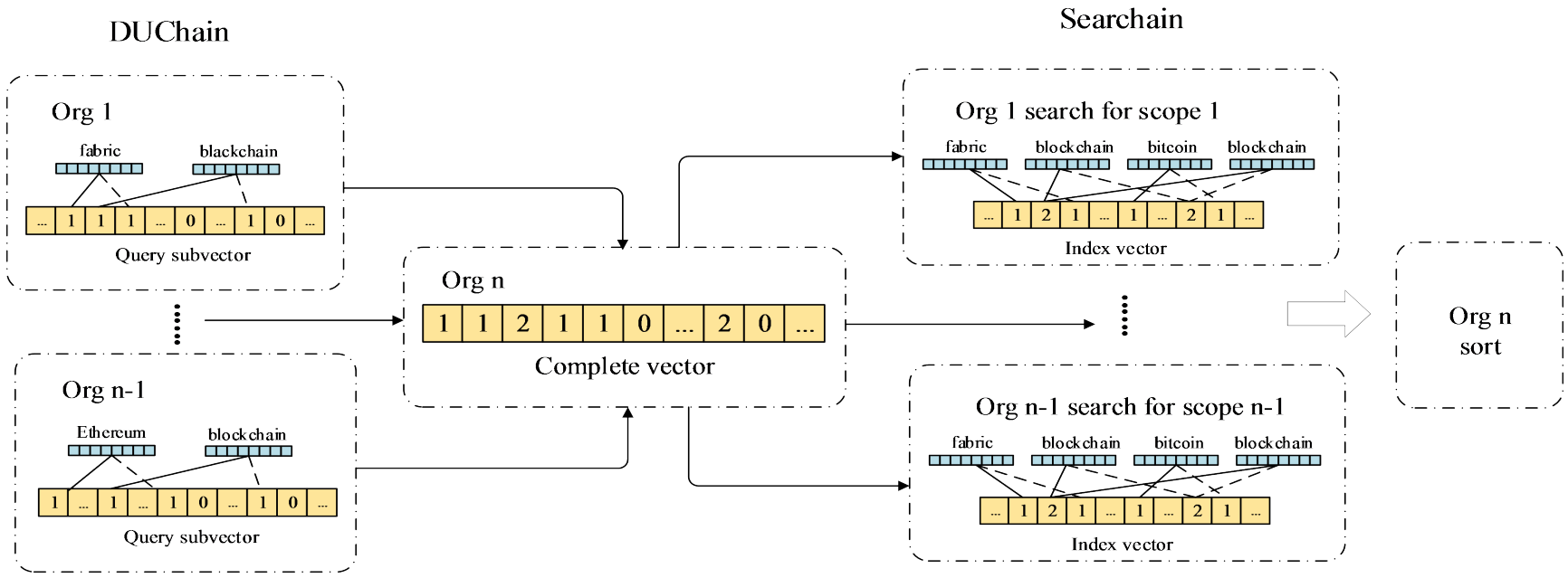
$$I'[t]^T Q'[t] + I''[t]^T Q''[t] = (I[t]/2 + r)Q[t] + (I[t]/2 - r)Q[t] = I[t]Q[t]$$

So,

$$I'^T \square Q' + I''^T \square Q'' = I^T Q$$

The search process

- The search process and an example diagram of index vector and query vector:



- **Experiments**

Experiments

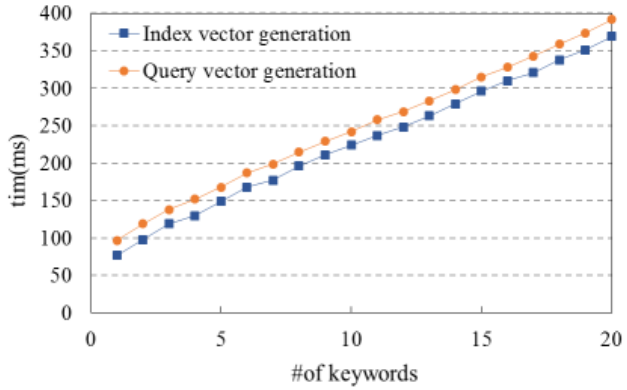
- We compared the function features of FMSM with the schemes [17], [19], [10] and [25] in terms of Multi-keyword search, Fuzzy search, Dynamic updates, Blockchain-based and Distributed-based:

Scheme	Multi-keyword search	Fuzzy search	Dynamic updates	Blockchain-based	Distributed-based
Wang[17]	√	√	√	×	×
Fu[19]	√	√	√	×	×
Liu[10]	√	×	×	×	√
Chen[25]	×	×	√	√	√
FMSM	√	√	√	√	√

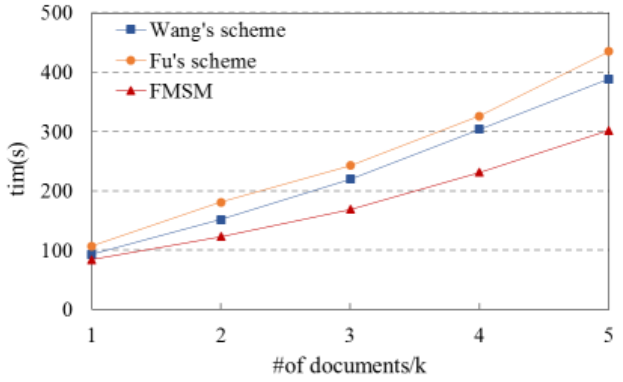
[10] Xueqiao Liu, Guomin Yang, Willy Susilo, Joseph Tonien, Ximeng Liu, and Jian Shen. 2020. Privacy-Preserving Multi-Keyword Searchable Encryption for Distributed Systems. IEEE Transactions on Parallel and Distributed Systems (TPDS) 32, 3 (2020), 561-574.
[17] Bing Wang, Shucheng Yu, Wenjing Lou, Thomas Y. Hou. 2014. Privacy-preserving Multi-keyword Fuzzy Search over Encrypted Data in the Cloud. In International Conference on Computer Communications (INFOCOM). 2112-2120.
[19] Zhangjie Fu, Xinle Wu, Chaowen Guan, Xingming Sun, and Kui Ren. 2016. Toward efficient multi-keyword fuzzy search over encrypted outsourced data with accuracy improvement. IEEE Transactions on Information Forensics and Security (TIFS) 11, 12 (2016), 2706-2716.
[25] Biwen Chen, Libing Wu, Huaqun Wang, Lu Zhou, and Debiao He. 2020. A Blockchain-Based Searchable Public-Key Encryption with Forward and Backward Privacy for Cloud-Assisted Vehicular Social Networks. IEEE Transactions on Vehicular Technology (TVT) 69, 6 (2020), 5813-5825.

Experiments

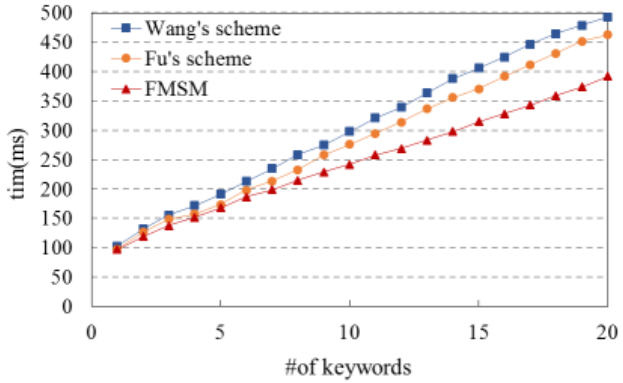
- Generation of secure index and trapdoor:



The variation of Index(Query) vector generation time with the number of keywords



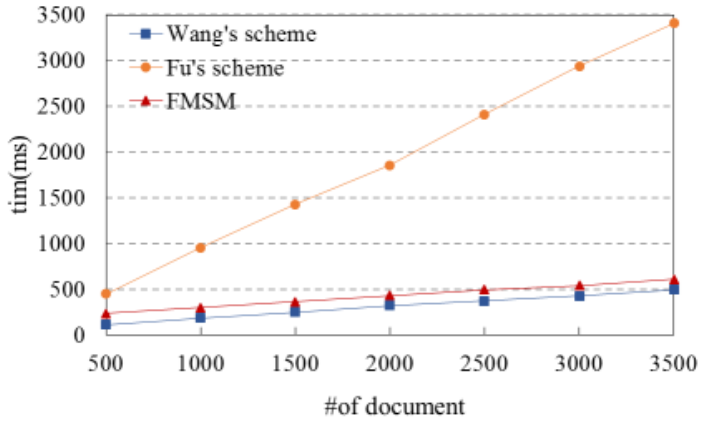
The variation of index generation time with the file size



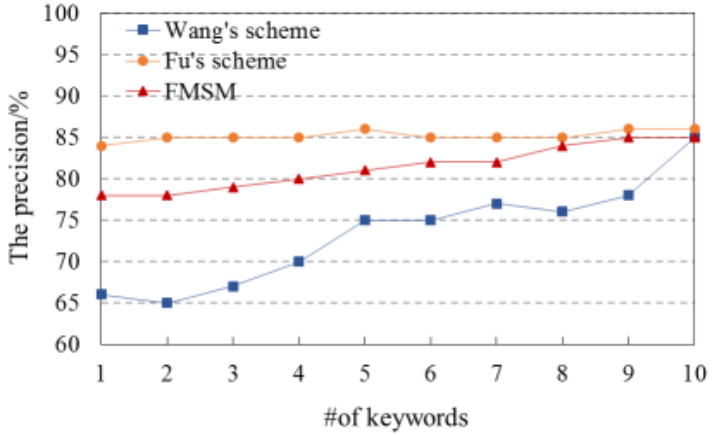
The variation of trapdoor generation time with the number of keywords

Experiments

- Search efficiency and accuracy:



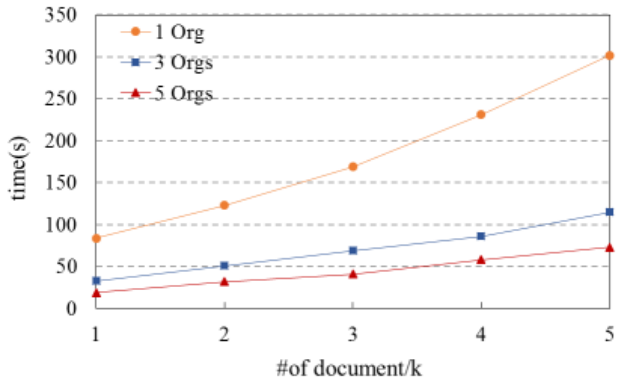
The variation of search time with the file size



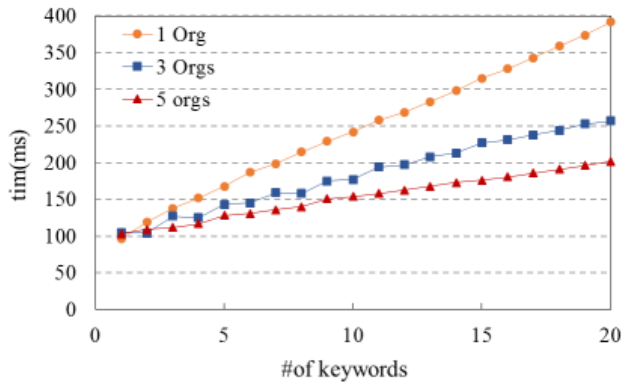
The variation of search accuracy with the number of keywords

Experiments

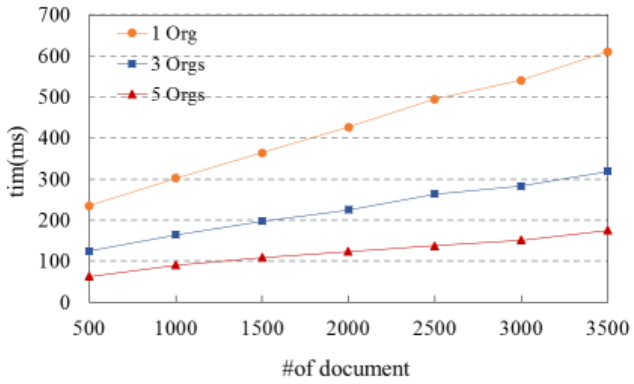
- Experiment analysis in a multi-node blockchain network environment:



The variation of the secure index generation time with the file size



The variation of the trapdoor generation time with the number of keywords



The variation of the search time with the file size

- **Conclusion**

Conclusion

In this paper, we proposed a Fuzzy Multi-keyword Search Scheme for Encrypted Cloud Data based on Multi-chain Network, namely FMSM, where we introduced blockchain and designed a multi-chain network to isolate participants and data, and achieve parallel processing of data. In FMSM, MinHash algorithm is introduced to reduce the dimensionality of the keyword set and Jaccard distance is used to calculate the vector similaity. In addition, the CBF is adopted as the index structure and the counts are indicated to the weights of the keywords. FMSM can efficiently support fuzzy multi-keyword search and dynamic updates for encrypted cloud data. Compared with the existing related schemes, FMSM achieved high security, reliability, search efficiency and accuracy.

Thank you for your attention