International Journal of Future Generation Communication and Networking
Vol. 13, No. 3s, (2020), pp. 286–295

Multi-Keyword Search for Banking Data using Block Chain

Krupal Kamble#1, Vicky Salve#2, Khilesh Patil#3, Atish Chavan#4, Geeta S. Navale#5

#Department of Computer Engineering, Sinhgad Institue of Technology and Science, Savitribai Phule Pune University, Pune.

wakeup.krupal@sinhgad.edu

vikassalve009@sinhgad.edu

khileshpatil16@sinhgad.edu

aatishchavan2@sinhgad.edu

gsnavale_sits@sinhgad.edu

Abstract

Due to the improvement of distributed storage, more information proprietors are slanted to re-appropriate their information to cloud administrations. For security concerns, sensitive data should be converted into hash blocks before outsourcing. In Blockchain technology, each page in a ledger of transactions forms a block. That block has an effect on the next block or page through cryptographic hashing. In other words, when a block is completed, it creates a unique secure code, which ties into the next page or block, creating a chain of blocks, or Blockchain. This paper discusses the related work carried out in this domain.

Keywords: Cloud, TF x IDF Model, Cryptography, Block chain technology.

I. INTRODUCTION

In recent years, cloud computing has been gaining much momentum in the IT sector that can be used to organize several IT resources, archiving and applications. Many IT sectors, banks and private companies are outsourcing their data to the cloud server. Many users can access and share information stored in the cloud regardless of locations. Cloud Service Providers (CSPs) can access confidential user data without authorization. The general approach of CSPs is to protect the confidentiality of the data in which the data is encrypted before outsourcing it to the cloud server and this will have a huge cost usability of data. In the secure search for encrypted data, data owners have outsourced their data to the cloud server in encrypted form for preserving your data. When the data user wants to search for any file, the data user sends a keyword request to the cloud server. Cloud servers then generate relevant results for the user of the data. The secure search on encrypted data not only reduces the processing and storage costs for secure keyword searches, but also supports searching for multiple keywords.

II. MOTIVATION

India is transferring into digital banking for that reason; banking sectors have to enhance the security of data which is stored in data centres and the cloud. Because in last few years there are lots of incidents happened where the data are stolen from data centre and cloud. Thinking on these issues, there is a need to enhance the security of data using Blockchain technology.

III. LITERATURE SURVEY

This section briefly summarizes the work carried out by other researchers.

Zhihua Xia et al. [1] a safe, effective and dynamic pursuit plot is proposed, which supports the precise multi watchword positioned search as well as the dynamic erasure and inclusion of archives. Proposed

ISSN: 2233-7857 IJFGCN

the security for protecting Multi-watchword Rank Searchable Encryption (MRSE) and set up a lot of exacting security necessities for such secure cloud information. TF x IDF for record development and question age was utilized in scrambled cloud information. Ravenous profundity first quest calculation was utilized for multi key hunt. Security is ensured by two danger models KNN. Among different multi keys semantically the effective standard of coordinating questions is utilized. Distributed computing has been considered as another model of big business IT framework, which can sort out enormous assets of processing, stockpiling and applications, and empower clients to appreciate universal, helpful and on request organized access to a common pool of configurable figuring assets with extraordinary effectiveness and negligible financial overhead.
However, Multi source frameworks give less security. Utilize open clouds these are untrusted mists.

Chang Liu et al. [2] developed a novel MDS (Multiple Data Source) model which securely stores data on the cloud.The secured data is made available for users using symmetric key encryption .They have used a symmetric encryption key for encryption. The same key used for decryption by the user. Their schemes outperform in security, storage and efficiently search data. However, the same key is used for encryption and decryption by both the parties which can be easily hacked and also a lengthy process to secure it. Social network data as well as its search ability in a secure manner. We see that current DSSE plots verifiably accept that the accessible file can be legitimately worked by a specific client. This presumption possibly bodes well when client information is very light weight and put away midway, which is anyway conflicting with numerous interpersonal organization applications. Consider an interpersonal organization, whose information is independently put away in a few server farms. Distributed computing has extraordinarily encouraged huge scope information re-appropriating because of its cost effectiveness, adaptability and numerous different favorable circumstances. Ensuing protection dangers power information proprietors to encode delicate information, henceforth making the redistributed information not, at this point accessible.
However, it is not always practical to make every data source have access to the online table. The usage of the online table has potential risks in real-world systems,

Jian Weng et al. [3] The human knowledge based publicly supporting comprises of three gatherings of jobs: requesters, laborers and a brought together publicly supporting framework A lot of laborers who are keen on this assignment contend and submit answers for the publicly supporting framework, while requesters will at that point select an appropriate arrangement (generally the first or the best one that explains the undertaking) and award the comparing laborers the prize. Publicly supporting includes getting work, data or conclusions from a huge gathering of individuals who present their information by means of the web, so they did it sagaciously by including Blockchain innovation for making sure about information. Blockchain is a kind of dispersed record for keeping up a lasting and carefully designed record of value-based information. A Blockchain capacities as a decentralized database that was overseen by PCs having a place with a distributed (P2P) arrangement utilizing keen agreement. Shrewd agreements were those applications that run precisely as modified with no chance of personal time, sensor boat, misrepresentation or outsider impedance. Before crowdsourcing there was no way to share tasks in the specific group. They were broadcasting the task on websites (for e.g. broadcasting information on up work, Wikipedia).
However, The Implementation of dynamic deletion and updating in documents. Data loss possible in this schema.

The problem was having some empty entries or corrupt data on crowdsourcing. Encrypted search by preserving privacy has been proposed by Xin Yao et al. [4] which developed the novel Multi Source Order Preserving Symmetric Encryption (MOPSE). The cloud server merged data indexes which were

ISSN: 2233-7857 IJFGCN

encrypted from multiple-data providers without revealing the content of the index. The data user can access data using query with the help of MOPSE. In addition to the MOPSE it had MOPSE+ to efficiently support the queries. The data providers and cloud servers were compromised through brute-force attack.
However, it is not supportable to a light-weight device like mobile; it had limited computation and memory resources.
Xin Yao et al. [5] which developed the novel Multi Source Order Preserving Symmetric Encryption (MOPSE). The cloud server merged data indexes which were encrypted from multiple-data providers without revealing the content of the index. The data user can access data using query with the help of MOPSE. In addition to the MOPSE it had MOPSE+ to efficiently support the queries. The data providers and cloud servers were compromised through brute-force attack. Again, it was not supportable to a light-weight device like mobile; it had limited computation and memory resources. The problem was having some empty entries or corrupt data on crowdsourcing. In this paper, we consider a multi-source CB-PHR framework in which different information suppliers, for example, medical clinics and doctors are approved by singular information proprietors to transfer their own wellbeing information to an untrusted open cloud. The wellbeing information are submitted in an encoded structure to guarantee information security, and every datum supplier likewise submits scrambled information lists to empower inquiries over the scrambled information. We propose a novel Multi-Source Order-Preserving Symmetric Encryption (MOPSE) plot whereby the cloud can combine the encoded information records from various information suppliers without realizing the list content.
However, Multi keywords are not used in this system.

B. Wang et al. [6] propose a novel multi-watchword fluffy hunt conspire by misusing the territory touchy hashing procedure. Our proposed conspire accomplishes fluffy coordination through algorithmic plan instead of extending the file record. It additionally kills the need of a predefined word reference and successfully bolsters various watchword fluffy ventures. The venue semantics play an important role in user check-in behaviour and modelled it using the heterogeneous user generated content. To the best of our knowledge, this is the first work that targets venue semantics using UGC. Different from the traditional geographical location representation, it represents the semantic information related to the locations. Different from the predefined location category representation, it is more flexible, and the UGC is readily available from social networks. In addition, it also takes geographical information into account by utilizing the venue context information.
However, the locations divide into 9 main categories, such as Arts & Entertainment, College & University, and so on. However, this location category also has some limitations

S. Pasupathi et al. [7] propose an effective and secure protection safeguarding approach for re-appropriated information of asset-obligated cell phones in distributed computing. Open key encryption calculation for scrambling the information and conjure positioned catchphrase search over the encoded information to recover the records from the cloud. We mean to accomplish a proficient framework for information encryption without giving up the protection of information. Further, our positioned catchphrase search incredibly improves the framework ease of use by empowering positioning dependent on pertinence score for query output, sends top most pertinent documents as opposed to sending all records back, and guarantees the fi le recovery exactness. Distributed computing, adaptable and flexible capacity and calculation assets are provisioned as estimated benefits through the Internet. Re-appropriating information administrations to the cloud permits associations to appreciate fiscal reserve funds, yet in addition improved nearby IT executives since cloud foundations are genuinely facilitated and kept up by the cloud suppliers. To limit the danger of information spillage to the cloud specialist organizations, information proprietors select to encode their touchy information, e.g.,

ISSN: 2233-7857 IJFGCN

wellbeing records, money related exchanges, before re-appropriating to the cloud, while holding the decoding keys to themselves and other approved clients.
However, Approach of result accuracy are an important performance metric. The Proposed scheme should find the results as accurate as possible and keep the accuracy within an acceptable range.

W. Sun et al. [8] proficient framework for information encryption without giving up the security of information. Further, our positioned catchphrase search extraordinarily improve the framework us capacity by empowering positioning dependent on pertinence score for output, send stop most applicable records as opposed to sending all documents back, and guarantees the record recovery accuracy.Thorough security and execution investigation, we demonstrate that our methodology is semantically secure and proficient. present a protection saving multi-catchphrase content pursuit (MTS) conspire with similitude based positioning to address this issue. To help multi-catchphrase search and query item positioning, we propose to assemble the inquiry record dependent on term recurrence and the vector space model with cosine comparability measure to accomplish higher output precision. ESPPA is a system that everything is client to look at by positioned catchphrase child scrambled information. Holding the security of the redistributed information of the proprietor while giving a way that permits a client to look effectively without the need of decoding the figure content.
However, significant drawback makes existing techniques unsuitable in Cloud computing as it greatly affects the system usability, rendering user search in experiences are very frustrating and system efficiency is very low.

G. W. Peters et al. [9] Distributed computing is another model of big business IT framework that empowers pervasive, advantageous, and on-request organize access to a mutual pool of configurable processing assets Due to the concentrated administration of flexible assets, all players in this rising X-as-an administration (XaaS) model, including the cloud supplier, application engineers, and end-clients, can receive rewards. presents a work which gives a graph of square chain development and its ability to upset the universe of dealing with a record through empowering overall money repayment, adroit agreements, automated keeping cash records and propelled assets. In such a way, they first give a brief blueprint of the middle pieces of this development, and likewise the second-age contract-based enhancements. To help multi-watchword search and item positioning, we propose to manufacture the pursuit record dependent on term recurrence and the vector space model with cosine closeness measure to accomplish higher query output precision.
However, the capability of the user to decrypt the received documents is a separate issue and is out of the scope of this paper.

L. Luu et al [10] presents a work which gives another circled understanding show for approval of less square chains called ELASTICO. ELASTICO is gainful in its framework messages and grants complex enemies of up to one-fourth of the total computational force. The higher the quantity of exchange squares chosen per unit time. ELASTICO is productive in its system messages and endures byzantine enemies of up to one-fourth of the all out computational force. Actually, ELASTICO consistently parcels or parallelism the mining system (safely) into littler councils, every one of which forms a disjoint arrangement of exchanges. The blockchain convention keeps up the circulated database in a decentralized system, accordingly meaning to tackle what we call the blockchain understanding issue. Reasonably, the issue is to permit a subjective huge system of a few processors to concede to the blockchain state (recognized by its cryptographic overview), under suspicion.
However, the agreement property in our problem is a relaxation of the original byzantine consensus problem. Which has its own scalability limitation? If the network size grows as we discussed.

ISSN: 2233-7857 IJFGCN

Table 1 presents the summary of the literature discussed above.

TABLE I SUMMARY OF LITERATURE REVIEW

Ref. No	Highlights	Observations
[1]	• Dynamic Searchable Symmetric Encryption (DSSE) is advanced . • Cryptographic primitive addressing the above issue, which maintains efficient keyword search over dynamic encrypted data without disclosing much information to the storage provider.	• Use a centralized cloud data centre.
[2]	• Proposed "Greedy Depth-first Search" algorithm to provide an efficient multi-keyword ranked search.	• Implement dynamic deletion and updating in documents • Data loss possible in this schema.
[3]	• Use the decentralized framework for Crowd-sourcing systems.	• Multi-keyword not used in this system.
[4]	• It helped to study how the patient data is stored on a cloud server via various data providers with his/her permission**.**	• Multi source systems provide less security. • Use a public cloud; these are untrusted clouds.
[5]	● In proposed system a novel multi-keyword fuzzy search scheme by exploiting the locality- sensitive hashing technique.	● The scheme achieves fuzzy matching through algorithmic design rather than expanding the index file.
[6]	● It helped to build the search index based on term frequency and the vector space model with cosine similarity measures to achieve higher search result accuracy.	● Used tree-based index structure and various adaptation methods for the multi-dimensional algorithm.
[7]	● It stores the encrypted data on the cloud and provides the search engine to retrieve files from the cloud. ● The main aim is to achieve an efficient system for a data encryption without sacrificing the privacy of data.	● It uses symmetric encryption.
[8]	Developed MTS (Multi Keyword Search) for searching data on the cloud.	● Instead of MTS use TF x IDF to search data.
[9]	● Provides a concept of Block chain Technology and how to use it in the banking sector to provide higher security.	● Used block chain instead of encryption and decryption algorithm.
[10]	● It proposes a distributed agreement protocol for permission less Block chain.	● It gives permission to empty nodes.

After going through observations listed in Table I, the proposed problem statement can be written as follows –

"To build and implement a multi keyword search on banking data using Blockchain technology."

The following objectives help to achieve the above-mentioned problem statement-

ISSN: 2233-7857 IJFGCN

- To provide the security for important data using Blockchain.
- To provide a multi keyword search over encrypted data.
- To provide the mediator for cloud data.
- To enable the cloud servers to perform a secure search without knowing any sensitive data.

IV. PROPOSED SYSTEM

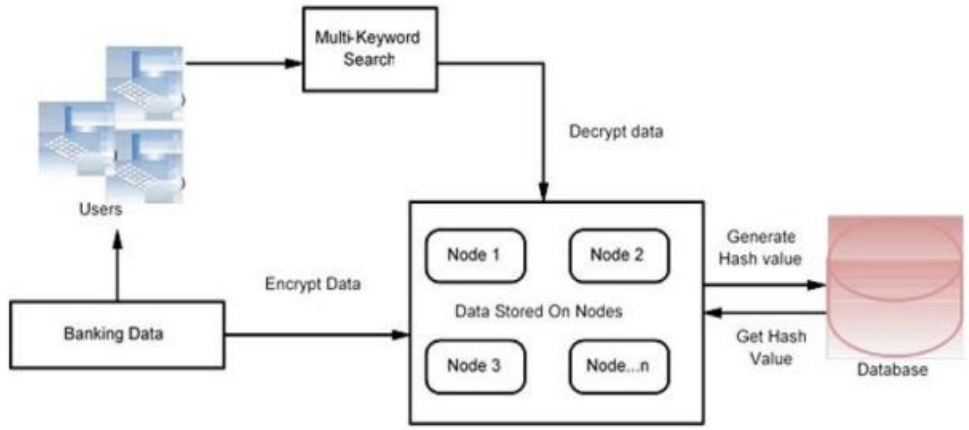

Figure.1: System Architecture

Figure.1 shows the architecture of the proposed system. Each data owner uploads data files on the cloud and encrypts the data with a corresponding key using blockchain. To implement both security preservation and efficiency searches, an efficient multi-keyword search scheme is proposed. In this system, the cloud server is allowed to effectively merge multiple encrypted indexes, and securely perform the multi-keyword search without revealing the data owner's sensitive information, neither data files nor the queries.

V. RESULTS

In experimental setup, the system shows file uploading time and file downloading time using the AES algorithm.

Sr.No	File Size(Kb)	Time(ms)
1	10351	226
2	17541	500
3	8500	140

ISSN: 2233-7857 IJFGCN

Table1: File Uploading Time and Size

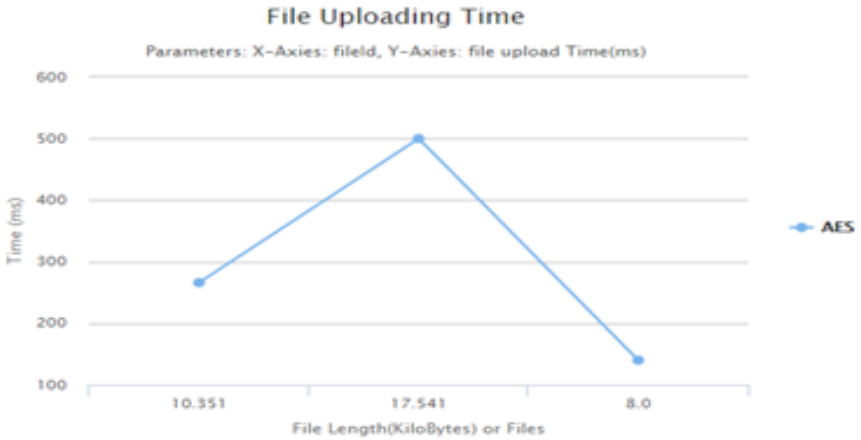

Figure.2a:File uploading time

Sr.No	File Size(Kb)	Time(Sec)
1	10351	22
2	25000	30
3	8000	10

Table1: File downloading Time and Size

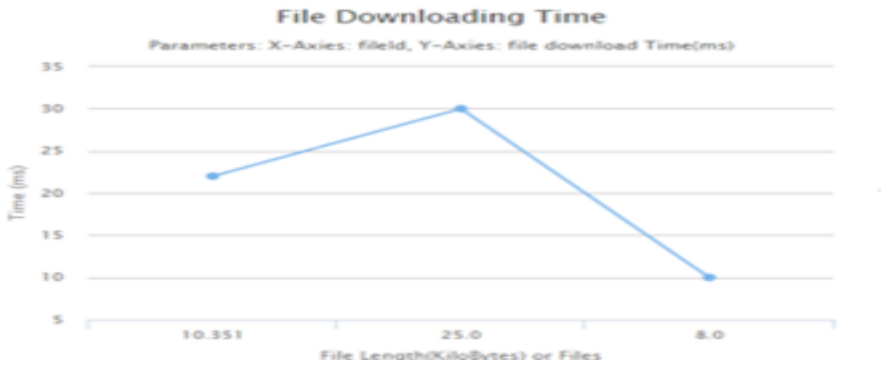

Figure.2b:File downloading time

ISSN: 2233-7857 IJFGCN

We developed a search engine using the TF-IDF model to search data from the cloud. In Figure.3a shown the GUI of search engine.After search keywords we got the list of files which shows Figure.3b

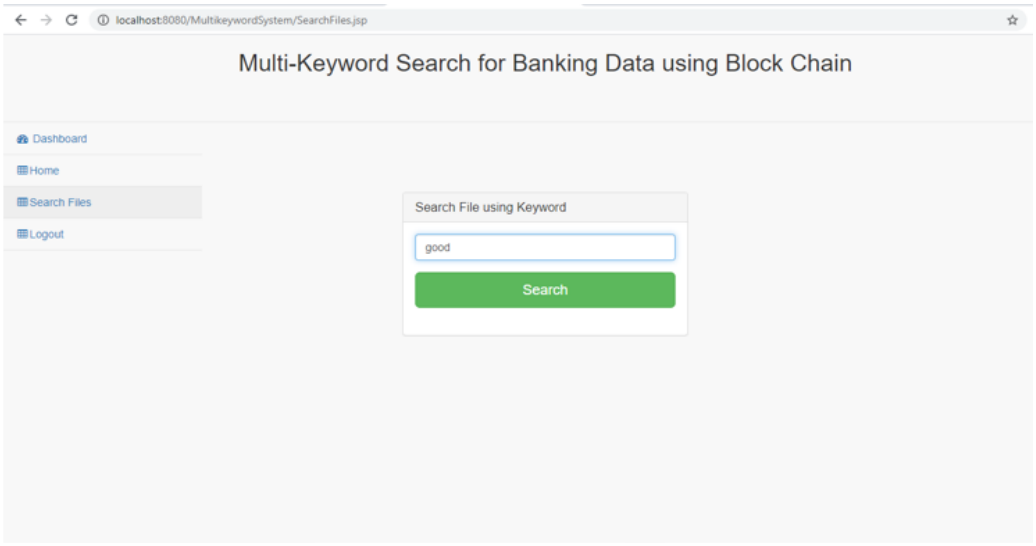

Figure.3a:Search Engine GUI

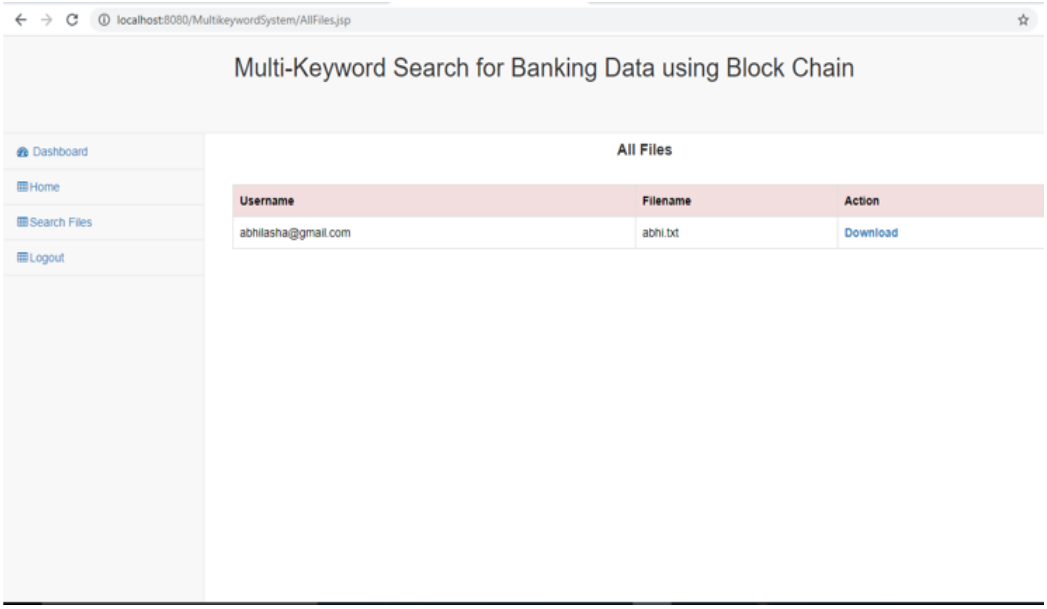

Figure.3b:Search Output

VI. CONCLUSION

In the proposed system data of the banking sector needs to secure. The proposed system uses blockchain concept and key-based cryptographic technique. The data is converted into a Hash node and stores on the cloud. Data validates other copies by running a hashing technique, and then compares the data stored in the Blockchain, any interference with the data will be quickly found, because the original Hash Tables have millions of nodes. The proposed system works on storing data of banking and supports multi-keyword search using TF x IDF. This system will work on consensus mechanisms while adding data in

ISSN: 2233-7857 IJFGCN

Blockchain. All the schemes discussed in introduction are limited to the single owner model. Therefore, this document proposes a multi-owner model to overcome the limitations of the previous methods, in which the encrypted data are found archived by multiple data owners and simultaneously data owners remain online to generate hatches. Different data owners share several secret keys to decrypt the secret data with different keys.

ACKNOWLEDGEMENT

The authors would like to thank the researchers as well as publishers for making their resources available. The authors are thankful to the authorities of Savitribai Phule University of Pune and concern members of Sinhgad Institute of Technology and Science, Pune for their constant guidelines and support. The authors also thank the college authorities for providing the required infrastructure and support. The authors are also thankful to the reviewers for their valuable suggestions.

REFERENCES

[1] Zhihua Xia, Xinhui Wang, Xingming Sun and Qian Wang, "A Secure and Dynamic Multi-Keyword Ranked Search Scheme over Encrypted Cloud Data", IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS, VOL. 27, NO. 2, FEBRUARY 2016.

[2] Tianyue Peng, Yaping Lin and Xin Yao, "An Efficient Ranked Multi-Keyword search for Multiple Data Owners over Encrypted Cloud Data", IEEE TRANSACTIONS ON CLOUD COMPUTING, VOL. 18, NO. 2 MARCH 2018.

[3] Ming Li, Jian Weng, Anjia Yang, Wei Lu,Yue Zhang, Lin Hou, JiaNan Liu, Yang Xiang, Robert H. Deng, "CrowdBC: A Blockchainbased Decentralized Framework IEEE Transactions on Parallel and Distributed Systems" (Volume: 30 , Issue: 6 , June 1 2019).

[4] Xin Yao, Yaping Lin and Qin Liu, "Privacy-preserving Search over Encrypted Personal Health Records in Multi-Source Cloud".

[5] X. Wang, Y. L. Zhao, L. Nie, Y. Gao, W. Nie, Z. J. Zha, and T. S. Chua, "Semantic-based location recommendation with multimodal venue semantics", IEEE Transactions on Multimedia, vol. 17, no. 3, pp. 409-419, Mar. 2015.

[6] B. Wang, S. Yu, W. Lou, Y. Hou, "Privacy-preserving multi-keyword fuzzy search over encrypted data in the cloud," in: INFOCOM'14, Toronto, Canada, 2014.

[7] S. Pasupuleti, S. Ramalingam, R. Buyya, "An efficient and secure privacy-preserving approach for outsourced data of resource constrained mobile devices in cloud computing," J NETW COMPUT APPL., vol. 64, pp. 12 – 22, 2016.

[8] W. Sun, B. Wang, N. Cao, H. Li, W. Lou, Y. Hou, H. Li, "cy presePrivarving multi-keyword text search in the cloud supporting similarity based ranking," IEEE T Parall Distr., vol. 25, no. 11, pp. 3025 – 3035, 2014.

[9] G. W. Peters and E. Panayi, "Understanding modern banking ledgers through blockchain technologies: Future of transaction processing and smart contracts on the internet of money," in Banking Beyond Banks and Money. New York, NY, USA: Springer-Verlag, 2016, pp. 239–278.

[10] L. Luu et al., "A secure sharding protocol for open blockchains," Proc. ACM SIGSAC Conf. Comput. Commun. Security, 2016, pp. 17–30.

ISSN: 2233-7857 IJFGCN