ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 2, 419-424 (2020)

www.astesj.com

ASTESJ

ISSN: 2415-6698

A Comprehensive Study of Privacy Preserving Techniques in Cloud Computing Environment

Bayan O Al-Amri, Mohammed A. AlZain, Jehad Al-Amri, Mohammed Baz, Mehedi Masud*

College of Computers and Information Technology, Taif University, Saudi Arabia

ARTICLE INFO

Article history:
Received: 06 December, 2019
Accepted: 04 March, 2020
Online: 04 April, 2020

Keywords:
Cloud computing, Multi-clouds, cloud storage, Privacy-preserving, Data privacy, multi-key, deep learning

ABSTRACT

The huge growth in cloud storage utilization over the past years has made a big demand for an advanced technique and strong tools to make services even more practical and secure. Data privacy in cloud computing has become one of the biggest concerns for both individuals and organizations which adds more pressure on cloud service providers to gain more trust. This paper surveys various privacy-preserving techniques in cloud computing fields, and addresses the highlights and tools used in each technique with an explanation of what and why these tools were used. This work aims to focus on the most innovative and strongest techniques that researchers has figured and tested so far.

1. Introduction

Cloud computing as defined by NIST [1] "a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction".

Over the years, cloud services have expanded more and more to catch the interests and demands of the users with less cost features of its services [2]- [5]. Users like normal individuals, or big expanding companies, or even governments, are all interested in the modern magic of safeguarding knowledge and information (AKA data). Though, perfection is nearly a wish in real life as well as in digital world, it seems that there are still a lot of work to do to reach the 100% of safety[5]-[7]. But aside from wishes, work and knowledge should be invested to prevent the most damage possible for the sake of integrity and privacy – preserving of data stored in clouds all over the network [8]-[10].

In cloud environment, data must be kept secure and save. If harm is done on the service provided to users regarding any kind of application serving a demand. Such as quick accessing from any place in the world, data storing, security measures against attacks, and many more of other privileges brought to users of cloud services [2] [11] [12] . There exist a lot of third-party cloud computing service and support providers of data, administration for processes, and much more of various needed services in the field. Their services yet are not just profitable, but offers even more of abilities in farm out data to cloud platform [13] [14] Cloud computing systems permit users to execute calculations on a massive quantity of data without building a whole groundwork from scratch [15] [16]. Yet, sending user's data to the cloud server in the form of a plaintext may endanger it to complete disclosure, which is a huge fail in privacy preserving condition [16]. Therefore, in aim to protect those data there should be some procedures and techniques to accomplish this goal. One of them is the mechanism of hiding data [17]-[19] by using specific calculations and algorithms.

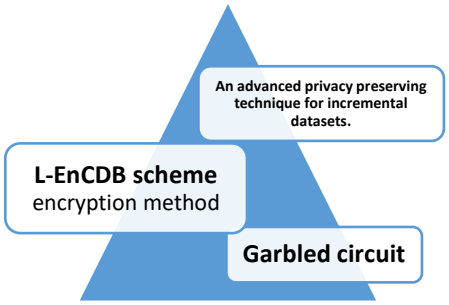

Figure 1: privacy preserving techniques.

2. Major Privacy Preserving Techniques

In this section the paper discusses three current privacy preservation techniques, which are most intriguing.

* Mehedi Masud, College of Computers and Information Technology, Taif University, Saudi ArabiaAddress, mmasud@tu.edu.sa

2.1. L-EnCDB Scheme

L-EnCDB is a novel lightweight encryption mechanism for databases[20]. This encryption method is using a format preserving encryption (FPE) scheme proposed with the L-EnCDB [20]. Secret and confidential data are generally encrypted before publishing to clouds for security necessities that results in a huge demand in practical operational databases. This is considered as a challenging mission; therefore, the authors proposed this technique claimed to success the challenges by addressing the following results:

- The technique supports practical SQL-based queries[20].
- Maintaining database structure [20].

After a detailed examination it turns out that this proposed technique is proven to be fairly practical and secure within the current security model [20]. The traditional methods of applying security to data used to be by encrypting data with techniques such as public key encryption, and symmetric key encryption [18] [21] [17, 22] [23], yet they change data types which makes it difficult to use them in database applications for several operations in SQL such as queries [20]. But still, all these data are categorized as private assets, along with information extracted through queries, so they cannot be unprotected if traditional encryption techniques do not meet the desired measures. This gives data owners a mission to both secure data and send them in an changeable form [20].

This is why L-EnCDB scheme was suggested. It proposes a trivial encrypted database method signified by L-EncDB, this method promises to achieve the necessities discussed beforehand, which are securing data without changing their format in a database to perform practically all kinds of operations needed in database applications [20]. L-EncDB can be successfully executed by using (FPE) technique and character string scheme, this way a data format can be preserved while in cipher text form [20]. In the proposed L-EncDB scheme, a trivial FPE scheme is proposed to support its trivial context for privacy preserving data in a database [20].

Note that the use of FPE enhancing security in databases in an obvious manner, its mission is to create a cipher text that is not changeable from plain text, thus the researchers developed a new FPE method, it basically has three algorithms, which they are, setup, encryption, decryption [20].

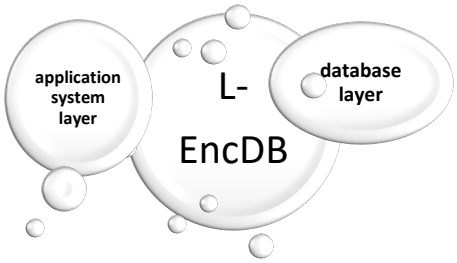

Figure 2: L-EncDB layers [20]

2.1.1 L-EncDB layers

There are two layers in this system, the first layer is the application system layer where several encryption methods (FPE, FQE, OPE) are used in several SQL queries, and the interface installed in the database will perform SQL analysis, encrypt, and create cipher texts, while the second layer is a database layer which is responsible for permitting developers to only perform SQL related functions, and also offer data services [20].

Security concepts:

SQL analysis interface is installed at the application layer or user's side of the system. In this system there expected to be a way to secure the key used in this encryption method, two kinds of adversaries might be colluded with the system's security:

- Those who target the database, with an attack to encrypted data and a database structure.
- Adversaries who target both the database and system's applications.

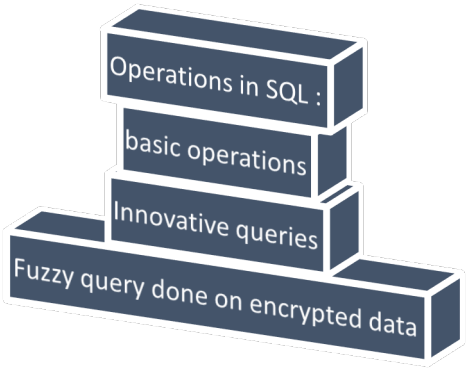

Figure 3: The three SQL operations on cloud[20].

Adversaries can gain entrée to interfaces or manipulate SQL applications in many ways, either by targeting the whole arrangement of the database, or by changing, building sentences in both ways of encryption and decryption. L-EncDB is believed to be more practical supporting fuzzy queries in SQL, and it offers a lighter and easier way to perform operations over data and sharing data securely to cloud providers. In conclusion, L-Enc-DB is suggested as an innovative technique that assures security and privacy preserving measures for data on cloud databases, capable of applying security over SQL application and various operations related to SQL in general, furthermore, with the ability of keeping data type and construction after applying encryption operations on them.

Along with the above advantages, it guarantees privacy preserving for queried data from the database, which means securing valuable data when linking the database to the cloud. This technique has been experimented and resulted that it is effectively secure when demonstrated over a huge database to achieve a stronger enhanced privacy preserving operations, as the founders of this advanced technique has viewed their scheme to prove a privacy preserving queries done on encrypted NoSQL database, thus for the sake of making the technique extra efficient it will be going under more examinations in the future for evaluations over cipher text, and will provide more effort to enhance privacy preserving querying from database along with a stronger publishing method supported by their method .

2.2 A Privacy Preserving Technique for Incremental Datasets:

Cloud Computing (CC) is known of being the service that offers data storage in a huge and powerful base between interconnected data servers and parties[5] [24], in which users of the service can share their data for storage or operations[25]. For the time being, there has been many cases of attacks on this simple service usage of sharing and storing. Therefore, the protection of data became a needed condition in the field of CC privacy. To fulfill this approach, there has been a lot of techniques proposed to protect data, some of them involve encryption techniques[18], such as FHE, but it does not serve well in case of incremental data sets due to its large cypher text. There exist other techniques which their concept is basically about anonymizing data sets to produce a secure and privacy preserving solution for shared data on clouds. But still, it's not achieving the paramount protection over the growth of datasets on CC, the spreading of enormous data capacity athwart several storage points bounds the privacy preservation. In this new anonymization technique, it is suggested to achieve a stronger security for huge data usage over several and increased datasets on cloud computing. The ability of privacy preservation and enhanced privacy necessities is verified over evaluation.

Many encryption techniques have been advanced and put under operating situations to maintain the security and preserve privacy for important data, yet they weren't useful enough due to some weaknesses and high cost.

Preserving privacy has other techniques which are the anonymization techniques such as generalization , anonymized data k anonymity, the anatomization, L-diversity .The incremental anonymization technique is basically about dividing anonymized data into several small blocks then be stored in the cloud, the anonymization is based on the level of K . The novel anonymized data will be initialized to allow the addition of more data to the cloud. Though, the new advanced anonymization technique is not built upon anonymization algorithms only, but along with storage and other operations, it preserves privacy of cloud data. Privacy necessities are maintained by an inception in the K-anonymity model.

Two major targets are reached:

1- An enhanced privacy preserving in cloud computing is brought successfully by the new anonymization technique, even though the utility of data storage in clouds is increased
2- In the case of performance overhead increased, data's privacy will be preserved without using another anonymization algorithm. Operation's time and data storage will be preserved tool.

How it works?

D* is the anonymized dataset of D

Anonymization level: K

B: blocks of data.

Both D and D* are installed in the cloud.

B is added to the cloud as:

$$B = \{b_1, b_2, \dots\dots, b_n\}$$

Then it will be added to D several times.

After that, the following $(D + b_i)^*$is generated, the target is to avoid performing a complete anonymization to $D + b_1 + b_2 + \cdots + b_n$ every time new blocks of b_i

Is added. Why? Because the operation is complicated resulting this procedure to be impractical and costly.

How to anonymize data without repeatition?

When the new data (b) are generalized upon K-old(b*), then added to previuosly added data D*. We check K-anonymous status and see if it is voilated and the anonymized datasets are over-generalized, then comparing and testing K-new to K-old:

- If K-old is minor than K-new? Secialization is made on sets with largest .
- K-old = K-new? Data is exported.
- Uneven generalization cases: carried out upon K-old.

Aouthors in their work [13] represented an algorithmatic code of the advanced anonymization technique. Noting that implementation time is depending on blocks numbers, while manifestation randomness is resulting from data variation. In summary, the new incremental anonymization technique should reach the operational goal efficiently plus overpowering performance overhead. Accordingly, the advantages of performig the new incremental anonymization technique is shown in decreasing the severe escalation in time and cost compared to the current anonymization techniques. Typically, clouds capacity of data is huge. It is revealed that the new anonymization technique can broadly enhance the privacy preservation on incremental datasets compared to recent techniques . This advanced incremenral anonymization technique is proven to be uneffected by the deviationK, and also proven to succesfully preserve privacy and confidentiality demmands

2.3 Garbled circuit

In the environment in which users trust their private data to third party servers AKA Clouds, there became a noticeable growth in using this type of data storage.

The need to preserve privacy has increased a lot recently, and it's not only urgent for companies and governments for big data, also to keep individual's information in their mobile devices away from disclosure as well, its keeping all data of all kinds and amounts protected. A portable device user shares data to a garbled circuit to perform calculations, these calculations are executed by the others server giving back the outcome of garbled calculations, this is showing the technique's way of procedures and measures to secure data and preserve privacy, this will guarantee the privacy

even if the calculating server did not gain data from all servers in the circuit. This technique includes the innovative use of the secured PRG of Blum et al. which allows the user to practically gain back the outputs of the computation along with assuring that it's been done accurately by the evaluator. Both the server and the user's processes will be analyzed by the system, to provide a privacy-preserving mechanism to user's device. Then the time of proceeding and constructing the above operations will be dignified to estimate the garbled circuit for multiple servers, proving the practicability of the secure cloud computing for mobile systems[26].

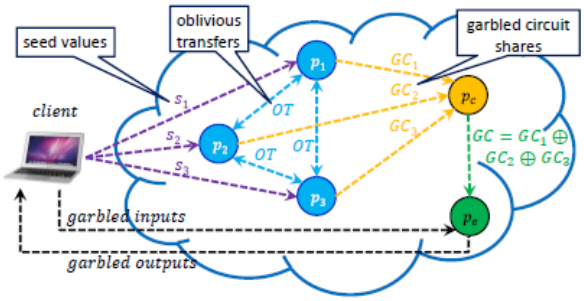

Figure4: secure cloud computing model with a mobile client using (2+2)=5 servers { P1,P2,P3,Pc,Pe} [26]

To explain the figure above, here are the details and steps as shown in figure 5:

1. Client sends seed values s_1, s_2, s_3 to p_1, p_2, p_3 respectively.
2. p_1, p_2, p_3 interact with one another to construct their shares of the garbled circuit, GC.
3. p_1, p_2, p_3 send their shares GC_1, GC_2, GC_3, respectively, to p_c.
4. p_c computes $GC = GC_1 \oplus GC_2 \oplus GC_3$, and sends it to p_e.
5. Client generates garbled inputs, and sends them to p_e.
6. p_e evaluates GC, and sends the garbled outputs to the client.

Figure 5: Explanation and Steps for GC [26]

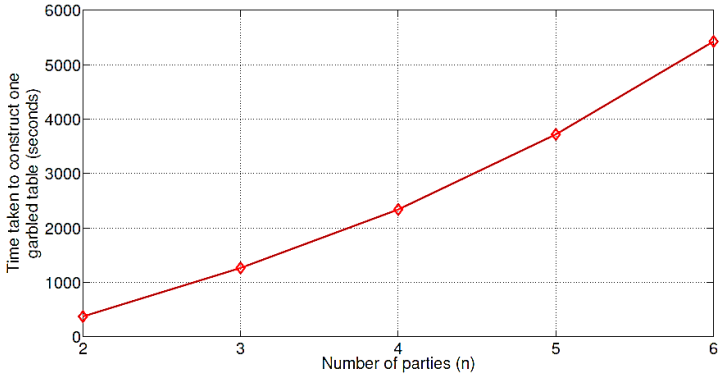

Figure 6: Consumed Time for Building the GC[26].

Approach highlights:

- In the proposed preserving of privacy solution of this approach, researchers compared their scheme with Gentry's FHE scheme , noting that it only uses one server ,but in comparison with their solution they prove it is more useful and secure explaining that with every garbled significance with a size of (nk+1) bits, for every input and output chain, the user simply interactions O(nk) bits with the server Pe.
- The consumed time in the process of building and checking the Garbled Circuit:

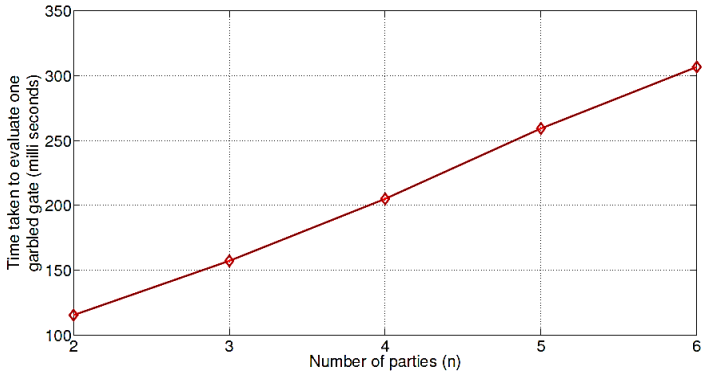

Figure 7: Consumed Time for Checking GC[26]

Fig. 6 displays the time consumed to build a single garbled compilation as a function of n. note that GC can be built in a corresponding way at every gate, and in this way, it can shrink the consumed time for building circuits. Fig. 7 displays the stage of checking a single garbled gate as a function of n. it's been shown that calculation is ominously faster than building, in case it's been done offline.

GCs can feely complete the requested operations when they already have been pre-operated and checked for a quicker response time for users.

Because of the unfeasibility of FHE schemes, due to the large cipher text, there has been an alternative solution using Yao's garbled circuits , for secure computations, it's been used with the multiparty computations with some other computations. In this scheme of multiparty computations, private inputs are held by several parties to attain the outputs of operations. However, in this advanced GC scheme, which is considered as a secure cloud computing system, users can only attain operation outputs from holding the preserved inputs in garbled method. Nevertheless, this upgraded GC scheme has been proven to match the necessities of preserved cloud servers by engaging several public cloud servers to operate the tasks of building and checking garbled circuits.

- By using this advanced GC, the user will be able to validate if in instance some unauthorized party has actually checked the GC, all that is effectively processed without engaging FHE encryption.
- Mobile user is the only one who has the authorization to deliver inputs to the secure computing model and attains outputs of the operations in garbled method.

- User's privacy of data inputs and outputs of the operations will be preserved despite the possibility of an evolving occurred between the checking server and a cloud exploiting in the process of constructing the GC.

This technique suggested an innovatively secure and confirmable cloud computing using several servers, the technique associates the protocol of et al.Goldreich (the secure multi-party computation protocol). And Beaver et al.'s GC design, along with the PRG method of Blum et al. Mobile user's privacy of data inputs and outputs of the operations will be preserved despite the possibility of an evolving occurred between the checking server and a cloud exploiting in the process of constructing the GC. A false checker will be discovered by not delivering values without the operations demanded, the system will perform an analysis for both sides, server and user of this system.

2.4 Information-Centric Approach:

Authors in [27] proposed a system using information-centric approach which makes cloud data self-intelligent. In this approach usage policy is used to encrypt and package cloud data. The data access mechanism consults its policy, create a virtualization environment, and assesses the trustworthiness of the data environment (using Trusted Computing). The architecture is shown in the Figure 8. The system is suitable in a corporate setting, in which database containing sensitive information need to be protected against external administrators, service providers, and local administrators. The system allows via machine readable rights expressions depth control over information that is allotted to a particular user.

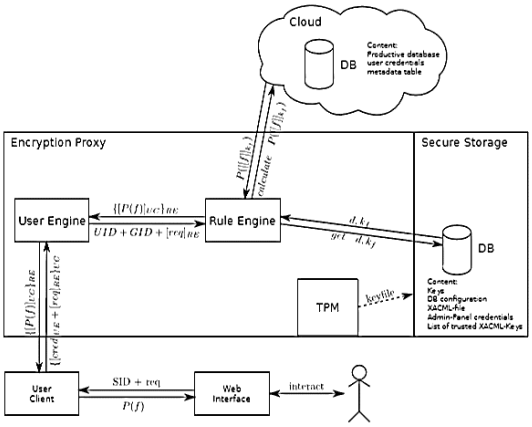

Figure 8: Information-centric approach [27]

2.5 Secure data-exchange in a cloud-based collaborative health care environment:

The authors in [28] presents a secure Cloud-based big data framework for collaborative healthcare service providers to efficiently store and manage large-scale health data. Each cloud data source exchange data using pair-wise communication for user queries by creating an on-the-fly data-exchange session. It provides a platform for sharing or exchanging health data residing in multiple clouds for the purpose of data analysis, decision making, and improving patients' treatment. The framework does not need a central third-party security mechanism (e.g., Public Key Infrastructure). Concerning pair-wise, on-the-fly data exchange, the authors presents a two-phase security protocol that uses pairing-based cryptography. Each cloud computes a secret session key dynamically by computing a pairing in an elliptic curve.

2.6 Privacy preserving data storage technique in cloud computing:

Authors in [29] proposed data fragmentation for privacy preserving data storage and retrieval. The mechanism protects the data unavailable when an unauthorized person compromised to one of this data splits in multi-cloud architecture. A decomposition technique introduced in this paper proves that privacy preserving information storage in multi-cloud is achieved. Based on the experiments, redundant storage of information is extremely efficient for data availability.

3. Results

In summary, all the techniques discussed above, here are some points and remarks regarding every method and some noticeable highlights, addressed in the following table:

Table 1:Privacy-Preserving Techniques in Cloud Computing

Field of remarks / Tech Name	Type of Privacy Preserving	Cloud field	Tools
EnCDB scheme	Encryption	This technique is applied over data on cloud database	L-EncDB & FPE
An advanced privacy preserving technique for incremental datasets.	Data anonymization encryption method	Preserving incremental data on cloud	Anonymization algorithm
Garbled circuit	PRG encryption over Garbled circuits.	While outsourcing data between multiple cloud servers	PRG method & a secure multiparty computation protocol

In Table 1 we addressed the three privacy preserving techniques in terms of main differences, tools and types of preserving techniques practiced in each one of them, as in the EnCDB scheme, encryptions using a format preserving encryption (FPE) scheme proposed with it is the method of privacy preserving here while it is an anonymization encryption method in the second technique using an advanced anonymization algorithm, the garbled circuit uses a PRG encryption method over the circuits.

The purposes and fields to apply each method differs in terms of the timing and practicability of performing them whether it is while outsourcing data between multiple cloud servers in the

Garbled circuit technique or where it is applied over data on cloud database in the EnCDB scheme, or it is for the sake of preserving incremental data on cloud in the second technique we addressed in this paper.

4. Conclusion

This paper describes the importance of data security and why we need privacy preserving techniques in cloud computing. The paper discusses different surveyed privacy preserving techniques proposed by researchers. Summary of the three important techniques are discussed. The cloud computing field is improving and offering several solutions to serve this goal privacy preserving. We believe that this paper will help whom ever has the need to improve further techniques in the privacy preserving field, to enhance a better yet stronger approaches.

References:

[1] (NIST), http://www.nist.gov/itl/cloud/.

[2] AlZain, M.A., et al. Cloud computing security: from single to multi-clouds. in 2012 45th Hawaii International Conference on System Sciences. 2012: IEEE.

[3] AlZain, M.A., B. Soh, and E. Pardede. Mcdb: using multi-clouds to ensure security in cloud computing. in 2011 IEEE Ninth International Conference on Dependable, Autonomic and Secure Computing. 2011: IEEE.

[4] AlZain, M.A., B. Soh, and E. Pardede, A new model to ensure security in cloud computing services. Journal of Service Science Research, 2012. 4(1): p. 49-70.

[5] AlZain, M.A., B. Soh, and E. Pardede, A survey on data security issues in cloud computing: From single to multi-clouds. Journal of Software, 2013. 8(5): p. 1068-1078.

[6] AlZain, M.A., B. Soh, and E. Pardede. A byzantine fault tolerance model for a multi-cloud computing. in 2013 IEEE 16Th International Conference On Computational Science And Engineering. 2013: IEEE.

[7] AlZain, M.A., B. Soh, and E. Pardede. A new approach using redundancy technique to improve security in cloud computing. in Proceedings Title: 2012 International Conference on Cyber Security, Cyber Warfare and Digital Forensic (CyberSec). 2012: IEEE.

[8] AlZain, M.A., Data security, data management and performance evaluation in a multi-cloud computing model. 2014.

[9] AlZain, M.A., et al., Byzantine Fault-Tolerant Architecture in Cloud Data Management. International Journal of Knowledge Society Research (IJKSR), 2016. 7(3): p. 86-98.

[10] Samra, H.E., B. Soh, and M.A. Alzain. A Conceptual Model for an Intelligent Simulation-Based Learning Management System Using a Data Mining Agent in Clinical Skills Education. in 2016 4th International Conference on Enterprise Systems (ES). 2016: IEEE.

[11] AlZain, M.A., et al., Managing Multi-Cloud Data Dependability Faults, in Knowledge-Intensive Economies and Opportunities for Social, Organizational, and Technological Growth. 2019, IGI Global. p. 207-221.

[12] Subashini, S. and V. Kavitha, A survey on security issues in service delivery models of cloud computing. Journal of network and computer applications. 34(1): p. 1-11.

[13] Aldeen, Y.A.A.S., M. Salleh, and Y. Aljeroudi, An innovative privacy preserving technique for incremental datasets on cloud computing. Journal of biomedical informatics. 62: p. 107-116.

[14] Fox, A., et al., Above the clouds: A berkeley view of cloud computing. Dept. Electrical Eng. and Comput. Sciences, University of California, Berkeley, Rep. UCB/EECS, 2009. 28(13): p. 2009.

[15] JoSEP, A.D., et al., A view of cloud computing. Communications of the ACM. 53(4).

[16] Liu, Z., et al. Secure storage and fuzzy query over encrypted databases. in International Conference on Network and System Security: Springer.

[17] AlZain, M.A., Utilization of Double Random Phase Encoding for Securing Color Images. International Journal of Computer Applications, 2018. 975: p. 8887.

[18] Faragallah, O.S., et al., Investigation of Chaotic Image Encryption in Spatial and FrFT Domains for Cybersecurity Applications. IEEE Access, 2020.

[19] Faragallah, O.S., et al., Secure color image cryptosystem based on chaotic logistic in the FrFT domain. Multimedia Tools and Applications, 2019: p. 1-25.

[20] Li, J., et al., L-EncDB: A lightweight framework for privacy-preserving data queries in cloud computing. Knowledge-Based Systems. 79: p. 18-26.

[21] Sodhi, G.K., et al., Preserving Authenticity and Integrity of Distributed Networks through Novel Message Authentication Code. Indonesian Journal of Electrical Engineering and Computer Science, 2018. 12(3): p. 1297-1304.

[22] Faragallah, O.S., et al., Block-based optical color image encryption based on double random phase encoding. IEEE Access, 2018. 7: p. 4184-4194.

[23] AlZain, M.A., Efficient Image Cipher using 2D Logistic Mapping and Singular Value Decomposition. INTERNATIONAL JOURNAL OF ADVANCED COMPUTER SCIENCE AND APPLICATIONS, 2018. 9(11): p. 196-200.

[24] Alraddady, S., et al., Deployment of Fog Computing During Hajj Season: A Proposed Framework. Procedia Computer Science, 2019. 161: p. 1072-1079.

[25] Alzain, M.A. and E. Pardede. Using multi shares for ensuring privacy in database-as-a-service. in 2011 44th Hawaii International Conference on System Sciences. 2011: IEEE.

[26] Premnath, S.N. and Z.J. Haas, A practical, secure, and verifiable cloud computing for mobile systems. Procedia Computer Science. 34: p. 474-483.

[27] Greveler, U., B. Justus, and D. Loehr. A privacy preserving system for cloud computing. in 2011 IEEE 11th International Conference on Computer and Information Technology. 2011: IEEE.

[28] Masud, M. and M.S. Hossain, Secure data-exchange protocol in a cloud-based collaborative health care environment. Multimedia Tools and Applications. 77(9): p. 11121-11135.

[29] Kartheeban, K. and A.D. Murugan. Privacy preserving data storage technique in cloud computing. in 2017 IEEE International Conference on Intelligent Techniques in Control, Optimization and Signal Processing (INCOS): IEEE.