Storage and encryption file authentication for cloud-based data retrieval

Mustafa Qahtan Alsudani[1], Hassan Falah Fakhruldeen[1,2], Heba Abdul-Jaleel Al-Asady[2,3], Feryal Ibrahim Jabbar[4]

[1]Department of Computer Techniques Engineering, Faculty of Information Technology, Imam Ja'afar Al-Sadiq University, Baghdad, Iraq
[2]Department of Electrical Engineering, College of Engineering, University of Kufa, Kufa, Iraq
[3]Department of Computer Technical Engineering, College of Technical Engineering, The Islamic University, Najaf, Iraq
[4]Department of Information Technology (IT), Al-Mustaqbal University College, Babylon, Iraq

Article Info

Article history:

Received Nov 1, 2021
Revised Feb 13, 2022
Accepted Feb 25, 2022

Keywords:

Benaloh homomorphic encryption
Cloud computing
Homomorphic encryption algorithm

ABSTRACT

The amount of data that must be processed, stored, and modified rises as time passes. An enormous volume of data from a wide range of sources must be stored on a safe platform. Maintaining such a large volume of data on a single computer or hard drive is impracticable. As a result, the cloud is the ideal platform for storing any quantity of data. An advantage of storing data in the cloud is that it may be accessed at any time and from any device. However, the security of data stored in the cloud is a big concern. Because of this, despite the benefits, most users are reluctant to move their papers to the cloud. The data should be encrypted before sending it off to the cloud service provider to avoid this issue. It's a great way to increase the security of your papers. According to a new technique presented in the system, data may be searched across encrypted files without compromising the privacy and security of various data owners. Implementing the pallier homomorphic encryption method makes it possible to perform computations on encrypted data without decryption.

This is an open access article under the CC BY-SA license.

Corresponding Author:

Hassan Falah Fakhruldeen
Department of Computer Techniques Engineering, Faculty of Information Technology
Imam Ja'afar Al-Sadiq University
Al-wazireya Near the Ministry of Labour and Social Affairs, Baghdad, Iraq
Email: hassan.fakhruldeen@gmail.com

1. INTRODUCTION

The cloud is an expression that refers to reaching computers, information technology (IT) and software apps through a network link, often via wide area networking (WAN) or internet communication WAN data center access [1]. Another benefit of cloud computing; provisioning the service is also easier. You can utilize it quickly in many situations; rather than being constrained by physical geography, remote users can access cloud services from anywhere they have a connection. It is also classified into three parts, which apply to access to services or infrastructure: private, public, and hybrid [2]. Anyone who wants to purchase or rent services may be offered public cloud services. Private cloud services were created by businesses to be utilized by their staff and partners only by integrating the two-hybrid cloud services [3].

A service that may be readily delivered through a network link, commonly using the web or mobile applications, has become known as cloud computing [4]. Consider cloud web hosting services (Amazon or Rackspace), digital content services (Apple iTunes, Amazon, and Netflix), cloud storage services such as dropbox or google drive, email services (Gmail), or even contracts for housing or transportation services

(Airbnb or Uber) [5]. Business programs such as microsoft outlook, typically used on local networks or devices, move to cloud-based apps [2], [3]. Four entities are involved in this scheme data owner, administration server, cloud server, and data usage. The data owner wants to upload sensitive data files into the cloud [6]. Due to privacy concerns, files must be encrypted before uploading them into the cloud [4]. With the approval from the administration server, the file is to be split and encrypted, with each block utilizing a standard symmetric encryption algorithm [7]. Before encrypting the files, keywords are extracted from the file and encrypted, and forwarded to store in the cloud server. Extracted keywords are encrypted homomorphically [8]. These encrypted keyword indexes made the searching operation easier. When the data user needs to retrieve the files from the cloud of his interest, firstly user sends a request to the administration server. The request is in the form of multiple keywords given to the administration server [5], [6]. The data can be encrypted utilizing any standard symmetric encryption algorithm like advanced encryption standard (AES) or data encryption standard (DES) [9]. In this manuscript, the AES encryption algorithm is utilized to encrypt the data files by the data owners. By doing so, searching over EncDta files is a tedious operation [10]-[15]. Upon receiving the request from the data, the user administration server authenticates the user. The administration server encrypts the received keywords homomorphically and sends them to the cloud server if the user is authorized. Otherwise, discard the request [16]. The searching operation to retrieve the files is done by the cloud server. If the server found a match between the encrypted keywords stored with the requested keywords, the server obtains the file id, retrieves it in a combined form, and forwards it to the requested data user. But the received file is in an encrypted form [17]. To access the file in the readable form only through the secret key utilized by the data owner to encrypt it. So, the user sends a request to the data owner to access the key. If the user receives the key from the owner, they can download the file and utilize it [7]. Storing and retrieving a considerable amount of data in the cloud, which can be accessed anywhere, at any time, and from any device by data owners or others, leads to poor security, and privacy about that data. So, the research problem is data security in the cloud to increase the privacy of data in the cloud, which will be stored in encrypted form. Encrypted information (encrypted information) may be calculated without first decrypting it using homomorphic encryption (HE). Upon decryption, the measurement result is encrypted, just as it would be if the operations were performed on the unencrypted results, and the results are identical. Homomorphic structures of encryption are fundamentally malleable [18]. HE schemes have poorer authentication properties in terms of malleability than non-homomorphic schemes.

Fully homomorphic encryption (FHE) makes it possible to test arbitrary unbounded depth circuits and is the best notion of homomorphic encryption [8], [18]. The multiplicative depth of circuits is the most functional constraint in conducting computations over EncDta for most homomorphic encryption systems. In 1978, within a year of the Rivest–Shamir–Adleman (RSA) scheme being written, the problem of building an FHE scheme was first suggested. If a compromise remained was unknown for more than 30 years. Partial outcomes over time included the following schemes [19]–[23]: i) RSA cryptosystem (unlimited modular multiplication count), ii) cryptosystem ElGamal (unbounded number of modular multiplications), iii) cryptosystem Goldwasser-Micali (unlimited number of exclusives or activities), iv) Cryptosystem BHE (unlimited number of modular additions), v) Paillier cryptosystem (number of modular adds without limits), vi) Sander-Young-Yung device (logarithmic depth circuits were overcome after more than 20 years), vii) the cryptosystem of Boneh-Goh-Nissim (unlimited number of addition operations but at most one multiplication), viii) Ishai-Paskin cryptosystem (branching programs of polynomial-size) [10], [21]. A searchable protected index must be generated for all documents submitted to the cloud to search documents efficiently. Before splitting the files into blocks, the administration server (AS) extracts keywords from the file utilizing a rapid keyword extraction algorithm [24]–[26].

2. METHOD

Cloud is the most promising medium to store and retrieve a large amount of data, which can be accessed anywhere and at any time and from any device. Here, it benefits this property of the cloud, while the security of the data in the cloud is the problem with this. To increase the privacy of the data in the cloud, which is to be stored in an encrypted form. The previous work performed the searching operation over un EncDta, a simple task while the security becomes overruled. To overcome the problem with searching over EncDta, here homomorphic encryption algorithm (HEA) is utilized for keywords encrypting to be stored and to be searched. HEA is utilized because of the reason that it helps to implement operations over EncDta without being decrypted. Hence, it improves the security of the files stored in the cloud and makes searching easier. The proposed system consists of four entities: data owner, AS, cloud server, and data usage. The data owner is the person who uploads files to the cloud. The administration server is an intermediate server between the owner/user and the cloud server. Uploading and requesting files are only through the administration server. The data user-made request to access and update the files to the cloud server. Cloud servers ultimately store the files, perform the searching operation, and provide the requested files to the data user. Figure 1 demonstrates the architecture of the proposed system.

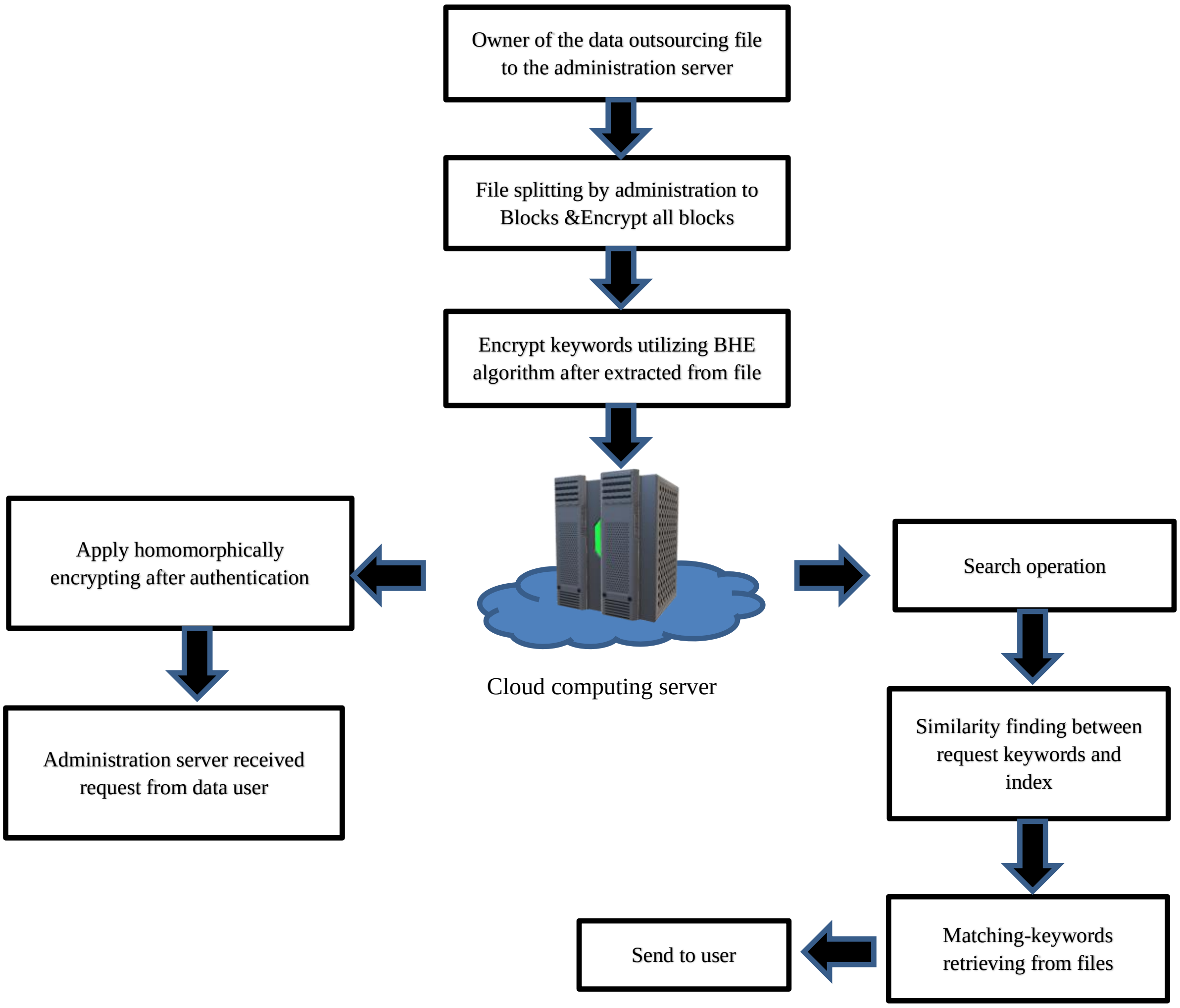

Owner of the data outsourcing file to the administration server
File splitting by administration to Blocks &Encrypt all blocks
Encrypt keywords utilizing BHE algorithm after extracted from file
Apply homomorphically encrypting after authentication
Administration server received request from data user
Cloud computing server
Search operation
Similarity finding between request keywords and index
Matching-keywords retrieving from files
Send to user

Figure 1. The system architecture

2.1. Processing unit side

Four separate organizations are involved in cloud storage: cloud server, system owner, service management, and user data. The data owner needs to register with the cloud-first. After that, the client needs to wait before the administration server accepts him. The cloud server hosts third-party data collection and retrieval facilities. The owner of the details should submit the records. Data must be encrypted until outsourcing since the submitted data can contain confidential data. The owner of the data transfers the file to AS. The file should be split into several blocks after obtaining permission from the AS. The block should be encrypted utilizing the standard symmetric encryption algorithm AES to forward the encrypted blocks to the cloud server.

2.2. Keyword extraction side

In this algorithm, the extraction process is done by four phases; preprocessing, word Co-occurrence graph, calculate word score, and keyword extraction.

2.2.1. Preprocessing

During the preprocessing step, the document text is partitioned into candidate keywords utilizing stop words and term delimiters. In a text, candidate keywords are word sequences that form content. A list of keywords for the candidate is generated by parsing the text. These delimiters are used to create an array of words from a single text string. At this point, a list is broken down into sequences of contiguous words and stop word locations (term delimiters). Whenever a group of words appears in a text simultaneously, they're regarded as potential keywords.

2.2.2. Word Co-occurrence graph

After the preprocessing stage, the word co-occurrence graph is drawn utilizing the candidate keywords from the previous stage. The X and Y-axis represent the candidate keywords. From the graph, the frequency of each word and how it is related to other words are identified.

2.2.3. Calculate word score

From the word co-occurrence graph, frequency and the degree of each word are obtained. The frequency of the word is the number of times that particular keyword occurs in the document. For example, word degree is the number of times a word appears within a lengthy list of possible search terms. Each potential keyword's score is calculated by dividing the degree by the frequency of the terms. By this method, the score value of all the candidate keywords is calculated. If the candidate keyword is extended, its score value is the sum of its member word scores.

2.2.4. Keyword extraction

Among the score value of all the candidate keywords, one-third of the words in the graph having the highest score is ultimately taken as the exact keywords utilized to create the searchable index.

2.3. Encryption side

The documents are encrypted using some encryption method before uploading into the global space, while homomorphic encryption is utilized for the stable index. The AES symmetric encryption algorithm is utilized to encrypt the document to be uploaded. There are two basic operations for specifying the encrypted text and the plaintext domain that homomorphic encryption performs. BHE is working over pure integers. It utilizes here the additive and multiplicative property of homomorphic encryption. Thus, BHE is applied to the keyword index. By utilizing BHE, computations can be performed over Encarta. Hence the keywords that need to be encrypted are firstly converted into integers. BHE is done is being as:

a. generating the key: select the most significant two primes p, q compute n=pq, n^2, and choose random integer g,

$$g = (1 + n) \tag{1}$$

Carmichael's Totient function conjecture

$$\lambda = lcm(p - 1, q - 1) \tag{2}$$

$$\varphi(n) = (p - 1, q - 1) \tag{3}$$

$$\mu = \varphi(n)^{-1}\, mod\, n^2 \tag{4}$$

Public key: (n, g) and private key: (p, q, λ)

b. encrypting: plaintext m, m<n *and* find a random r.

$$Ciphertext\ c = \ g^m . r^n mod\, n^2 \tag{5}$$

2.4. Searching operation

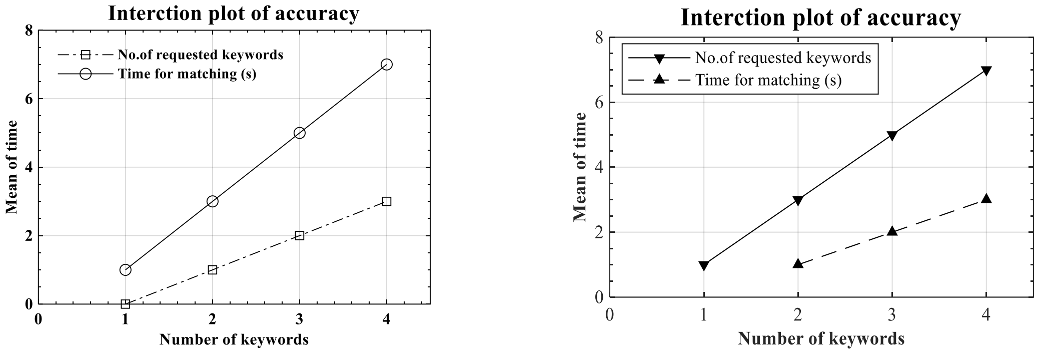
The cloud server does the search operation. Homomorphically encrypted keywords are stored in the cloud server when the data user requests the administration server by multiple keywords. After authentication, the authentication server encrypts the requested keywords homomorphically, and forwards them to the cloud server. In the cloud server, both the stored and requested keywords are in the integer form by applying BHE. Then the searching can be performed by subtracting both integer values, i.e., stored keyword value and requested keyword value [12], [19]. Search operation can be done between the ciphers is being as:

$$Difference, d = \frac{E_n[a,r]E_n[b,r]^{-1} - 1\, mod\, n^2}{n}\, mod\, n \tag{6}$$

If the subtraction result is zero, the ciphers are identical; otherwise, they are not. Through this method, exact keywords can be found. From the matched keywords, the corresponding file id can be obtained. If the matching file is found, combine the split blocks of that particular file and send it to the data user. The corresponding encrypted files are sent to the requested data users by the cloud server. At the same time, the secret key utilized for encrypting the file is required by the data used to download the contents of the file. For that purpose, the data user sends a request forwarded to both AS and the data owner to get permission to download the received file. The file can be downloaded only after getting permission from the owner and the AS. The data owner provides permission in the form of the secret key and AS again cross-checks the timestamps and the requested nature of the data user. If it was valid, then approve to download the file. The downloaded file can be updated by the data user with the permission of the corresponding data owner. The user sends an edit request to the corresponding data owner. The owner can view changes made by the user. If he approves, the user can make changes to the file. If the owner disagrees with the changes in the file's contents, then the owner can block the user from further operation to be done. Hence, the user no longer remains a data user in that cloud environment.

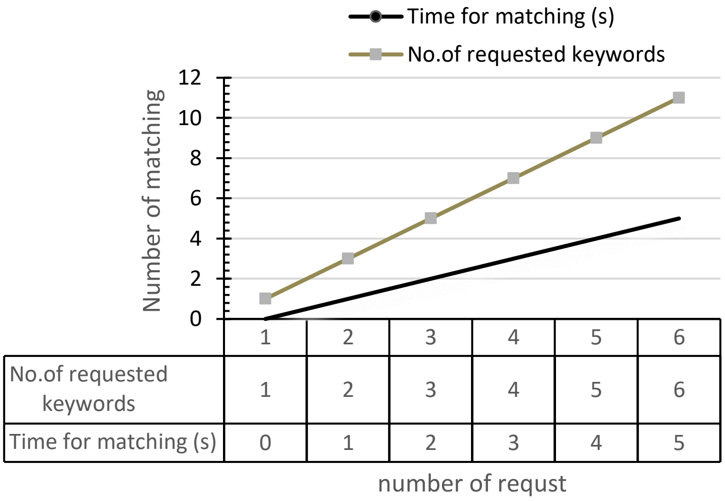
3. RESULTS AND DISCUSSION

The proposed system brings much better performance than the existing systems, which perform the search operation utilizing cosine similarity (MKSCS). In MKSCS, indexed keywords are obtained by utilizing the porter stemmer method and encrypted utilizing any standard encryption algorithm during outsourcing to the cloud. The matching operation is performed over unencrypted keywords instead of encrypted. The encrypted keywords are decrypted, perform the matching operation, and find suitable files on searching. The performance of the proposed system is evaluated as compared with the existing MKSCS. Co-occurrence statistical information-based keyword extraction is used for each metric's outcomes. Figures 2 concerning the number of keyword comparisons clearly show that the predictive performance typically improves as the number of keywords maintained in the dataset increases. Figures 2(a) and (b), candidate keyword no. vs. real keywords. Figure 3. no. of real keywords vs. significance. Figure 4. no of requests keywords vs. time for operation matching. The result of the performance evaluation is shown below:



(a) (b)

Figure 2. Candidate keyword no. vs. real keywords (a) main effect plot for accuracy and (b) main effect plot for time matching (s)



	1	2	3	4	5	6
No.of requested keywords	1	2	3	4	5	6
Time for matching (s)	0	1	2	3	4	5

Figure 3. No. of real keywords vs. significance

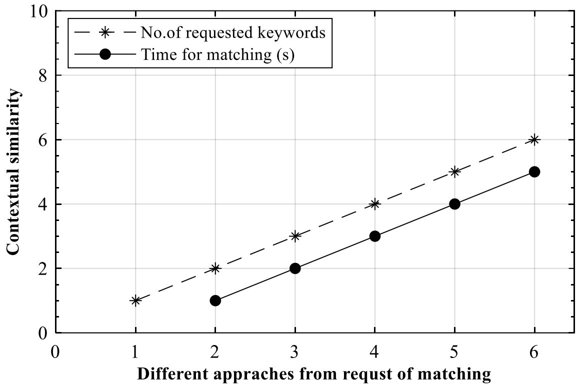


Figure 4. No of requests keywords vs. time for operation matching

4. CONCLUSION

Connection to global storage space over the internet is becoming popular every day. Security is the primary issue of storing data in a trusted third party. Deploying a safe multi-word search over encrypted cloud data is a big issue since the data is encrypted before it's sent out. Data owners will be encouraged to save their EncDta files in a global storage system, where they will be searchable and retrievable by many authorized users. Computing can be achieved without decryption by using HEA. As the server does not know the exact data to be processed by the data owner and requested by the data user, it helps to enhance data protection.

REFERENCES

[1] C. Wang, N. Cao, J. Li, K. Ren, and W. Lou, "Secure Ranked Keyword Search over Encrypted Cloud Data," *2010 IEEE 30th International Conference on Distributed Computing Systems*, 2010, pp. 253–262, doi: 10.1109/ICDCS.2010.34.

[2] J. Li, J. Li, X. Chen, C. Jia, and Z. Liu, "Efficient keyword search over encrypted data with fine-grained access control in hybrid cloud," in *International conference on network and system security*, vol. 7645, pp. 490–502, 2012, doi: 10.1007/978-3-642-34601-9_37.

[3] C. Liu, L. Zhu, L. Li, and Y. Tan, "Fuzzy keyword search on encrypted cloud storage data with small index," *2011 IEEE International Conference on Cloud Computing and Intelligence Systems*, 2011, pp. 269–273, doi: 10.1109/CCIS.2011.6045073.

[4] W. Sun, *et al.*, "Privacy-preserving multi-keyword text search in the cloud supporting similarity-based ranking," in *Proceedings of the 8th ACM SIGSAC symposium on Information, computer and communications security*, May 2013, pp. 71–82, doi: 10.1145/2484313.2484322.

[5] Z. Xu, W. Kang, R. Li, K. Yow, and C. -Z. Xu, "Efficient Multi-Keyword Ranked Query on Encrypted Data in the Cloud," *2012 IEEE 18th International Conference on Parallel and Distributed Systems*, 2012, pp. 244–251, doi: 10.1109/ICPADS.2012.42.

[6] C. Wang, S. S. M. Chow, Q. Wang, K. Ren, and W. Lou, "Privacy-Preserving Public Auditing for Secure Cloud Storage," in *IEEE Transactions on Computers*, vol. 62, no. 2, pp. 362–375, Feb. 2013, doi: 10.1109/TC.2011.245.

[7] D. X. Song, D. Wagner, and A. Perrig, "Practical techniques for searches on encrypted data," *Proceeding 2000 IEEE Symposium on Security and Privacy. S&P 2000*, 2000, pp. 44–55, doi: 10.1109/SECPRI.2000.848445.

[8] R. Curtmola, J. Garay, S. Kamara, and R. Ostrovsky, "Searchable symmetric encryption: improved definitions and efficient constructions," *Journal of Computer Security*, vol. 19, no. 5, pp. 895–934, 2011, doi: 10.3233/JCS-2011-0426.

[9] A. Chatterjee and I. Sengupta, "Searching and sorting of fully homomorphic encrypted data on cloud," *Cryptology ePrint Archive*, pp. 1–14, 2015.

[10] F. Baldimtsi and O. Ohrimenko, "Sorting and searching behind the curtain," in *International Conference on Financial Cryptography and Data Security*, Springer, Berlin, Heidelberg, January 2015, pp. 127–146, doi: 10.1007/978-3-662-47854-7_8.

[11] S. Rose, D. Engel, N. Cramer, and W. Cowley, "Automatic keyword extraction from individual documents," *Text mining: applications and theory*, vol. 1, pp. 1-20, 2010.

[12] T. Sridokmai and S. Prakancharoen, "The homomorphic other property of Paillier cryptosystem," *2015 International Conference on Science and Technology (TICST)*, 2015, pp. 356–359, doi: 10.1109/TICST.2015.7369385.

[13] S. W. Kareem, R. Z. Yousif, and S. M. J. Abdalwahid, "An approach for enhancing data confidentiality in hadoop," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 20, no. 3, pp. 1547–1555, December 2020, doi: 10.11591/ijeecs.v20.i3.pp1547-1555.

[14] Q. Shallal, Z. Hussien, and A. A. Abbood, "Method to implement K-NN machine learningto classify data privacy in IoT environment," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 20, no. 2, pp. 985–990, November 2020, doi: 10.11591/ijeecs.v20.i2.pp985-990.

[15] S. Shyla and S. Sujatha, "Efficient secure data retrieval on cloud using multi-stage authentication and optimized blowfish algorithm," *Journal of Ambient Intelligence and Humanized Computing*, pp. 1–13, 2021, doi: 10.1007/s12652-021-02893-8.

[16] S. Sharma and S. S. Rajput, "Dynamic Key Generation Based Data Retrieval in Cloud Environment," *International Journal of Modern Engineering & Management Research*, vol. 5, no. 4, pp. 23–28, December 2017.

[17] G. S. Mahmood, D. J. Huang, and B. A. Jaleel, "Data security protection in cloud using encryption and authentication," *Journal of Computational and Theoretical Nanoscience*, vol. 14, no. 4, pp. 1801–1804, April 2017, doi: 10.1166/jctn.2017.6508.

[18] S. Mudepalli, V. S. Rao, and R. K. Kumar, "An efficient data retrieval approach using blowfish encryption on cloud ciphertext retrieval in cloud computing," *2017 International Conference on Intelligent Computing and Control Systems (ICICCS)*, 2017, pp. 267–271, doi: 10.1109/ICCONS.2017.8250724.

[19] B. Bülbül and D. T. Altılar, "Privacy Preserving Data Retrieval on Data Clouds with Fully Homomorphic Encryption," *2019 4th International Conference on Computer Science and Engineering (UBMK)*, 2019, pp. 1–6, doi: 10.1109/UBMK.2019.8907057.

[20] N. H. Hussein, "Cloud-Based Efficient and Secure Scheme for Medical Images Storage and Sharing using ECC and SHA-3," *2019 2nd Scientific Conference of Computer Sciences (SCCS)*, 2019, pp. 109–115, doi: 10.1109/SCCS.2019.8852620.

[21] M. Raje and D. Mukhopadhyay, "Algorithm for Back-Up and Recovery of Data Stored on Cloud along with Authentication of the User," *2015 International Conference on Information Technology (ICIT)*, 2015, pp. 175–180, doi: 10.1109/ICIT.2015.16.

[22] H. B. Mahajan *et al.*, "Integration of Healthcare 4.0 and blockchain into secure cloud-based electronic health records systems," *Applied Nanoscience*, pp. 1-14, 2022, doi: 10.1007/s13204-021-02164-0.

[23] M. Q. Alsudani, S. H. A. Reflish, K. Moorthy, and M. M. Adnan, "A new hybrid teaching learning based Optimization-Extreme learning Machine model based Intrusion-Detection system," *Materials Today: Proceedings*, July 2021, doi: 10.1016/j.matpr.2021.07.015.

[24] S. Li, C. Xu, Y. Zhang, Y. Du and K. Chen, "Blockchain-Based Transparent Integrity Auditing and Encrypted Deduplication for Cloud Storage," in *IEEE Transactions on Services Computing*, doi: 10.1109/TSC.2022.3144430.

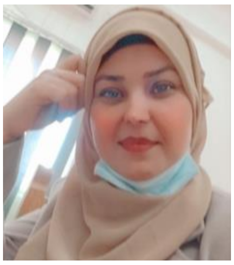
[25] H. A.-J. Al-Asady, H. F. Fakhruldeen, and M. Q. Alsudani, "Channel estimation of OFDM in c-band communication systems under different distribution conditions," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 23, no. 3, pp. 1778–1782, September 2021, doi: 10.11591/ijeecs.v23.i3.pp1778-1782.

[26] S. K. S. Raja, A. Sathya, S. Karthikeyan, and T. Janane, "Multi cloud-based secure privacy preservation of hospital data in cloud computing," *International Journal of Cloud Computing*, vol. 10, no. 1-2, pp. 101–111, March 2021.

BIOGRAPHIES OF AUTHORS

Mustafa Qahtan Alsudani received the B.Eng. degree in technical computer engineering from Al-Rafidain University College, Iraq, in 2010 and the M.S. and Ph.D. degrees in computer engineering techniques from National Aerospace University "KhAI," Kharkiv, Ukraine, in 2013 and 2018, respectively. Currently, he is an Associate Professor and Head of the department of computer engineering teachings, imam Jaafar Al-Sadiq University. His research interests include computer vision, system vulnerability, system security, cyber-attacks, wireless networks, and wireless communications. He can be contacted at email: alsudani.m.q@mail.com.

Hassan Falah Fakhruldeen received the B.Eng. degree in communications engineering from Al-Furat Al-Awsat Technical University, Iraq, in 2010 and the M.S. and Ph.D. degrees in electronics and communications engineering from Baghdad University, Iraq, in 2013 and 2020, respectively. Currently, he is an Associate Professor at the Department of Electrical Engineering, University of Kufa. His research interests include photonics, optics, optical fiber communications, nano-photonic devices, plasmonics devices, optical communications, optical fiber networks, plasmonic sensors, all-optical signal processing, 5G communications, communications transmission lines, signal transmission planning, wireless networks, wireless communications, and radio over fiber communications. He can be contacted at email: hassan.fakhruldeen@gmail.com.



Heba Abdul-Jaleel Al-Asady received the B.Eng. degree in electrical engineering from Al-Kufa University, Iraq, in 2010 and the M.Sc. and Ph.D. degrees in electronics and communications engineering from Babylon University, Iraq, in 2014 and 2022, respectively. Currently, she is an Associate Professor at the Department of computers Engineering at Islamic University. Her research interests include communication, electronics, encryption systems, cryptographic technologies, information systems, sensors, digital signal processing, 5G communications, transmission lines, signal transmission planning, wireless networks, and wireless communications. She can be contacted at email: en.he22@gmail.com.

Feryal Ibrahim Jabbar received a B.Eng. degree in electric power from Al-Furat Al-Awsat Technical University, Iraq, in 2008, in 2010, and the M.S. University of Mosul, Department of Electrical Engineering (2015), and PhD, (2021). Artificial Intelligent engineering from University Tun Hussein Onn Malaysia (UTHM), Department of KFEE. She is a teacher at the Department of IT, Al-Mustaqbal university college, Iraq. Her research interests include work in artificial intelligence applications and renewable energy fields. She can be contacted at email: faryal.ibrahim@mustaqbal-college.edu.iq.