Secure Blockchain Based Data Storage and Integrity Auditing in Cloud

Sumathi M[a], Rajkamal M[b], Dr B Gomathy[c], I Infant Raj[d], Dr. Sushma Jaiswal[e] and D. Swathi[f]

[a] Assistant professor, Department of Computer Science and Engineering,
K.Ramakrishnan College of Engineering, Samayapuram, Trichy.
[b]Application developer, IBM Bangalore
[c]Professor/CSE, KPR Institute of Engineering and Technology
[d]Assistant Professor, Department of Computer Science and Engineering,
K. Ramakrishnan College of Engineering Trichy
[e]Assistant Professor, Department of Computer Science & Information Technology (CSIT)
Guru Ghasidas Vishwavidyalaya, (A Central University), Koni, Bilaspur, (C.G.), India, 495009
[f]Assistant professor, Department of Computer Science and Engineering,
K.Ramakrishnan College of Engineering, Samayapuram trichy.

Article History: Received: 10 January 2021; Revised: 12 February 2021; Accepted: 27 March 2021; Published online: 20 April 2021

Abstract: A blockchain is a linear collection of data elements, where each data element is called as block. All blocks are linked to form a chain and secured using cryptographic hash function. Each block typically contains a hash pointer as a link to a previous block, a timestamp and transaction data. Public verification techniques enable users to outsource the data integrity verification to a dedicated third party auditor. To ensure the security in the case that the auditor is compromised, the users are required to audit the auditor's behaviors. Hence, the user data need to maintain in a secure way is an essential task along with data integrity. Blockchain provides such kind to secure data storage along with data integrity without compromising confidentiality. The double SHA provides a collision free secure storage than other technique. Hence Double SHA based block generation and verification is proposed in this work. In this work both data owner and public auditors are verified the integrity for providing highest accuracy. Experimental results show that the proposed technique provides better results and security to user data in an efficient manner is proved.

Keywords: Blockchain, Public Audit, Hash Value, Double SHA, Confidentiality and Integrity.

1. Introduction

Cloud computing provides the platform for data storage and accessing service resources from cloud through network. Data sharing as one of the most common features in cloud storage, allows a number of users to share their data with others. A potential method of solving this problem is to encrypt the whole shared file before sending it to the cloud. Generate the signature used to verify the integrity of this encrypted file, finally upload this encrypted file and its corresponding signatures to the cloud. Auditing helps to detect the unauthorized access and modifications with the help of third party auditor. Occasionally these service providers and public auditors also behave as adversaries. To overcome these issues, the blockchain technology is used financial and medical sectors [8].

Blockchain is a blooming technology which is used to store the data in a decentralized storage location. The key features of blockchain technology are decentralized storage, tamper proof record maintenance, immutability, enhanced security and records are maintained in distributed ledgers. Due these characteristics blockchain technology is used in many organizations like medical, financial and government organizations for maintaining a record without data loss [13]. The major issues of cloud storage technique are third party service provider and public auditor control. Occasionally these service providers and public auditors also behave as adversaries. To overcome these issues, the blockchain technology is used financial and medical sectors. In a blockchain technology, the centralized control is eliminated by the members are involved in a network. The tamper proofing record maintenance provides integrity to user data. Similarly, in cloud computing the public auditors verify the data access information in a frequently static time intervals. In a blockchain technique, data integrity is verified in a dynamic manner through tamper proof records [9].

Blockchain is created in two different modes like permission-less and permissioned. In a permission-less mode, everyone is possible to join in a network and able to create block based on puzzle solving process. The information which is stored in a permission-less blockchain is able to view by all members are involved in a network. There is no confidentiality is maintain in this network [10]. To overcome this issue, the permissioned blockchain is used for maintaining confidential information like medical and financial data. In a permissioned network, the data owner is having the full control over their network. Only the registered members are able to participate and access the information in this network. When compared to permission-less network, permissioned network provides confidentiality and integrity to user sensitive information. In a cloud storage technique, occasionally authorized user access is denied by cloud service providers [11]. This drawback is overcome through

blockchain based decentralized storage. The authorized users are able to access the information from a registered network with minimal effort. Hence, the data availability is achieved through blockchain technology.

The major drawback of blockchain technique is storage capacity such as the storage capacity of each block is 1MB only. When a data size is increased, the performance degradation will be occurred in a blockchain technique. Similarly, the user data is stored in each user distributed ledger. This storage technique increases data availability and consistency but, it will require large storage space utilization [12]. Hence, an alternate technique is required to utilize the blockchain technique in an efficient manner.

The remaining sections of this article are organized as follows. In section 2, the works that are related to blockchain technique based storage is discussed with its merits and issues. Section 3, our proposed system is discussed with system architecture and necessary algorithms in a detail manner. In section 4, the proposed technique experimental results are discussed with existing work. Finally in section 5, the proposed work is concluded with its future enhancements.

2. Related Works

In this section the works that are related to blockchain based data storage and public auditing techniques is discussed with its merits and issues. Xiaodong Yang et al. discussed about the blockchain based multi-replica and multi-cloud data with public auditing scheme. In this work, the issues that are related to third party public auditing scheme is discussed with multiple storage issues. To reduce storage overhead, the replicated data were [14]stored in different cloud storage locations. The third party service providers and public auditor'sinteractions are overcome through unpredictability of block storage in a blockchain. Through this process, the computational and communication overheads are reduced. In this technique, the replicated information is stored in a cloud storage leads to a security issues [1]. Li Yanna et al. proposed the privacy preserving cloud data auditing technique. In this work, the key updating and authenticator evolving mechanism is used key verification process. This process reduced the communication and computation cost while maintaining the desirable security. [15]But, the it's difficult to update cloud users secret auditing key [2].

ShenJian et al. used public auditing protocol with global and sampling blockless verification scheme for cloud data verification. This proposed protocol performed well in terms of [18][19]efficient dynamic support and reduced overhead. The major overhead of this technique was, increased the burden on data owners who are not equipped with sufficient computing resources [3]. ShenWenting et al, discussed the homomorphic invisible authenticator technique. This technique protects the privacy of authenticator and [20][21]supports the blockless verification. But, this technique infeasible to large scale computation and communication process [4]. ShenWenting et al. established the storage auditing scheme for group users. In this scheme, the user does not need to perform time-consuming decryption process. The major issue of this scheme was,key exposure should not be neglected [5]. Yu Yong et al. proposed the identity based [16] remote data integrity checking protocol. The RDIC protocol provided security against the malicious server. This protocol lacked with predominant governing body to enforce these standards [6][22].

TianHui et al. [24] used public auditing with random masking scheme. This random masking scheme effectively achieves secure auditing process. But, the computational costs for user revocations are relatively high in this technique [7][23].

Limitations of an existing works:

- Cannot resist a procrastinating auditor who may not perform the data integrity verification on schedule.
- Deviate from the original objective of public verification schemes.[17]
- It might be too late to recover the data loss or damage if the auditor procrastinates on the verification.
- The procrastinating auditor also cannot be detected in the public verification schemes even though malicious auditors can be detected there.

Problem Statement:

In public verification schemes, after data outsourcing, the user sets a verification period. Then the auditor verifies the outsourced data integrity at the corresponding time when receiving auditing request. Most public verification schemes are built on the public key infrastructure where the auditor needs to manage the user's certificate to choose the correct public key for verification.

Objective:

- To use blockchain technology, provides a tamper-proofing and distributed way to provide data security without a central authority.
- Third Party Authenticator (TPA) who verify the integrity group of data stored in the cloud whether the auditing proof is correct or not.

3. Proposed Work

The key idea of the proposed work is, the public verification technique is that the data owner splits the data into multiple blocks and computes a signature for each block using double SHA 256 algorithm. The hashed data blocks outsources along with the corresponding signatures to the cloud server. When the auditor verifies the data integrity, it chooses a random subset of all data blocks and sends the sampled blocks indexes to the cloud server. The cloud server responses with the corresponding proof, the auditor checks the integrity of challenged blocks by verifying the validity of the proof. Figure 1. shows the proposed system architecture.

This proposed blockchain based data storage restricts the behavior of third party auditors. Similarly the certificate-less public audit scheme increases efficiency of auditing process. This scheme provides data owner based access monitoring and tracking system. Hence, the data owners are dynamically monitor their data access in an efficient manner. Additionally, the proposed technique resists the collusion attack to TPA and CSP. Similarly, the proposed system allows the cloud server to prove that the outsourced data is well maintained. The communication and computation overhead should be as efficient as possible. This technique improves the public auditing and privacy preserving with the help of hashing approach. Fast auditing with higher performance protocols is used for better results. The working principle of the proposed system is discussed as follows.

- User Enrollment – Before going to create or access a block, the user has to be registered with the system for getting the authentication and authorization. In a basic authentication process, a user presents some credentials like user ID to prove that the user is the true owner of the user ID. Afterwards, the data owner could upload the file on cloud. Once the file is stored in a cloud, the file will be getting encrypted.

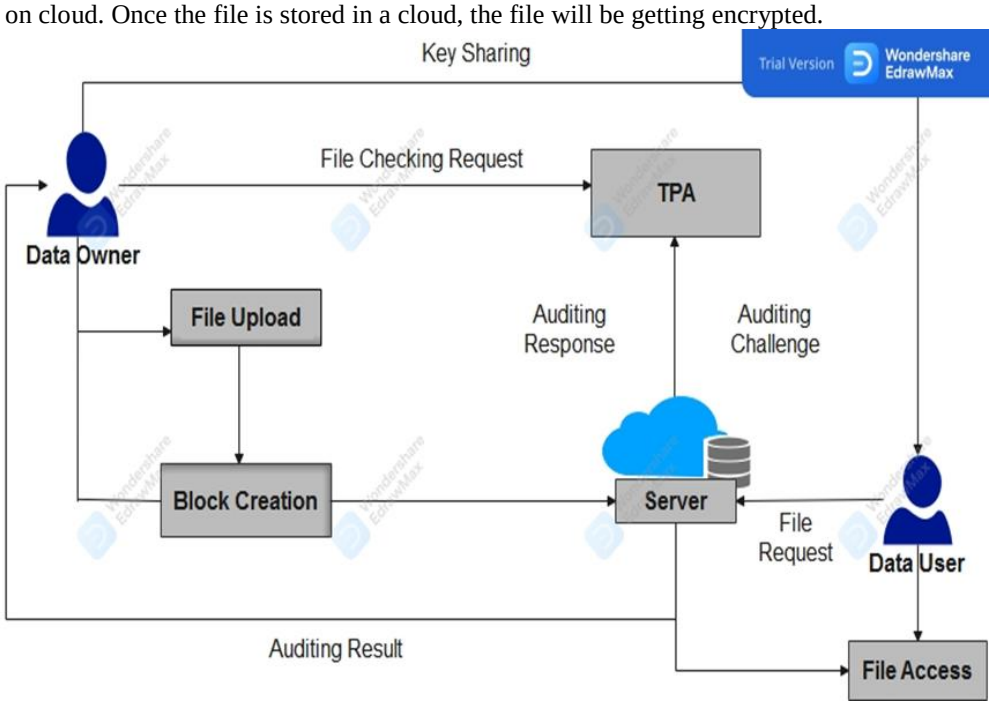

Figure 1. Proposed System Architecture

Data Block Creation –In blockchain based storage, the block is a digital representation of data is stored it. When a block of data is chained to the other blocks, its data can never be changed again. Blockchain technology functions are reliable for use in a hashing crypto method. Which helps create an adequate and strong hashing code and convert it from a bit of fixed size data to strings of character. Each transaction proposed in a blockchain is hashed together, and the hash pointers connect each block to the next block. Any changes in the blockchain transaction of hashing function will result in different hash string of character and affect all the involved blocks. Equation 1 is used for create a block based on data block, hash code of current block (HCB) and the hash code of the previous block (PBH).

Data block=Double SHA_256(Data+HCB+PBH) (1)

When compared to SHA_256, Double SHA_256 is collision free and it is suitable for hexa-decimal based block generation. Hence, double SHA_256 is preferred in for the block generation. In algorithm 1 the block generation process is discussed.

Blockchain Storage -The storage scheme of data uses blockchain based cloud storage technology to achieve safe storage and sharing.To create a local Cloud and provide priced abundant storage services. Data storage and access control are the main transactions in the medical blockchain.Once get space from cloud the users can upload to share data in the cloud.In this work, the cloud storage can be implementing with high secure using blockchain technology. Figure 2. Shows the blockchain generation with blocks.

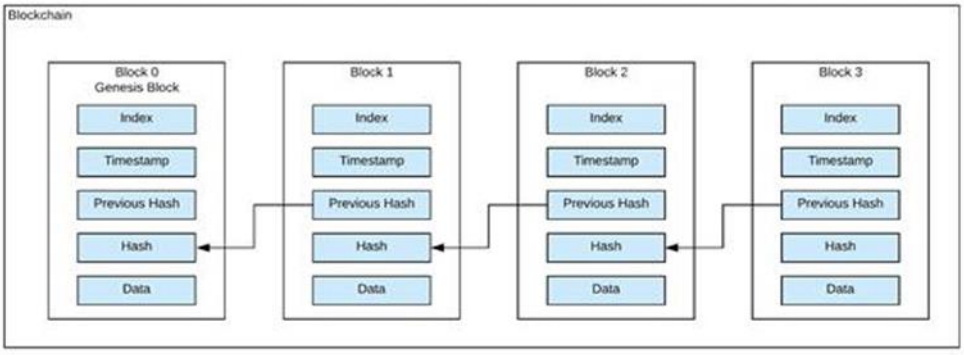

Figure 2 Blockchain Generation

• Data sharing - In the data sharing concept storage server is most important process.The storage data store the huge amount of data. This data is securely store in storage server. It also store encrypted data and key which used for data encryption. When the user requires his data, user requests to the storage server. There are two keys used for encryption and decryption purpose. Data sharing can be done in a secure manner. Figure 3. Shows the block encryption and decryption process.

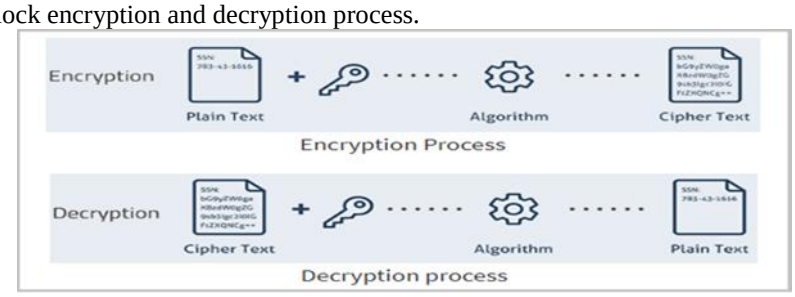

Figure 3. Block encryption and decryption

Algorithm 1: Block Generation and Verification

Input : Data, Double SHA_256, PBH

Output : Block of Data

Procedure

1. Login using the registered ID.
   a. Input → File + PBHV + CBHV
2. Read input and split into an array of chunks with 512 characters.
3. Use double SHA to generate hash code

$$Data\ block = Double\ SHA\_256(Data + HCB + PBH)$$

4. To generate block with hash code of data, HCB and PBH
5. Give audit request to the TPA.
6. TPA forwards the request to the cloud server and input as the index of block to be audited.
7. Cloud server audits the specified file and displays output.

   If file found

   Output → file found

   Else if file is modified

   Output → File found and modified

   Else output → File not found

• Data Auditing - TPA works for the user. It feeds back the verification results to the user and the cloud server, and detects the data corruption as soon as possible. The communication between TPA and other entities is authenticated. Upon receiving the challenging message, the cloud server computes the

corresponding proof. TPA checks the validity of the proof to verify the data integrity. If the checking fails, TPA informs the user that the data may be corrupted. Figure 4 shows the data auditing process.

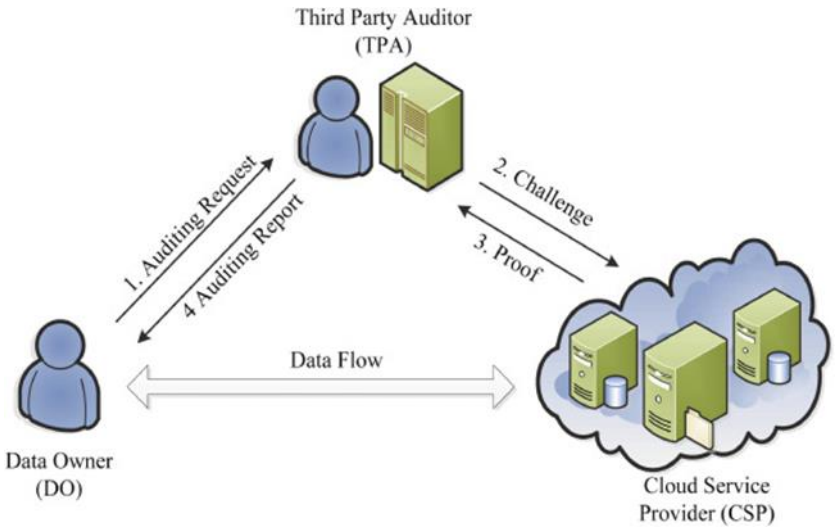

Figure 4.Cloud data auditing process

1. Data owner send an auditing request.
2. The third party auditor send a this request as a challenge to cloud service provider.
3. Based on challenge the cloud service provider sends a proof to third party auditor.
4. Now, the third party auditor sends an auditing report to data owner.

4. Experimental results

The proposed technique is implemented in java language with Intel Pentium system. The Double SHA 256 is used for hash code generation along with data, HCB, PCB. Figure 5 shows the encryption process along with double hash code generation. This process includes the id, owner name, file information, file name, file size, keys which are used for encryption and hash code generation, hash code 1 and 2.

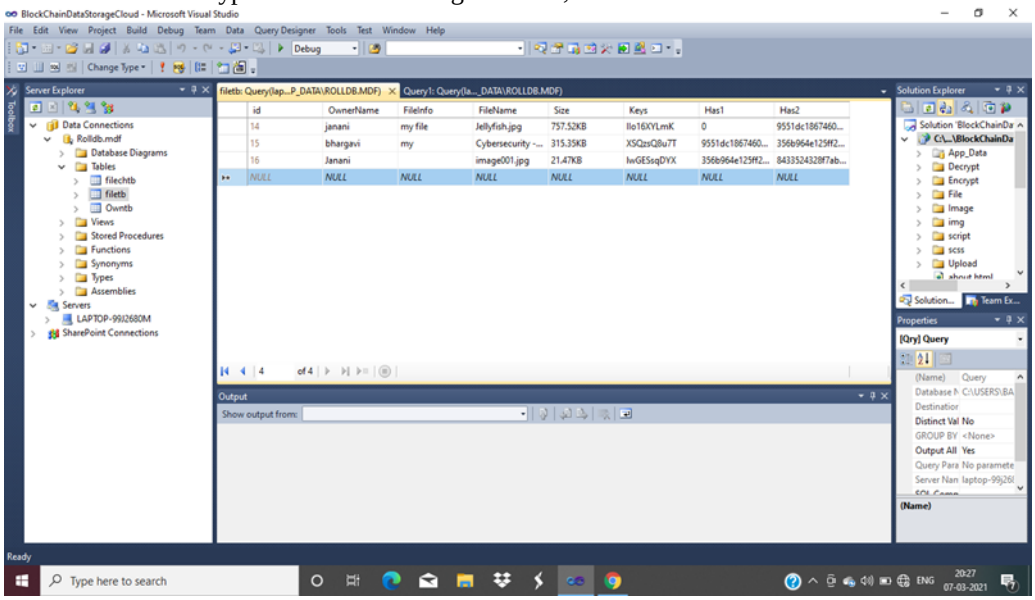

Figure 5. Data encryption and Hash Code generation

Figure 6.shows the block request along with verification details. It includes the id, owner name, file information, file name, status and action which is taken based on integrity of the block. Through the status, the data owner and third party public auditor knows the file found or not. Similarly, the file found with modification is also identified through it.

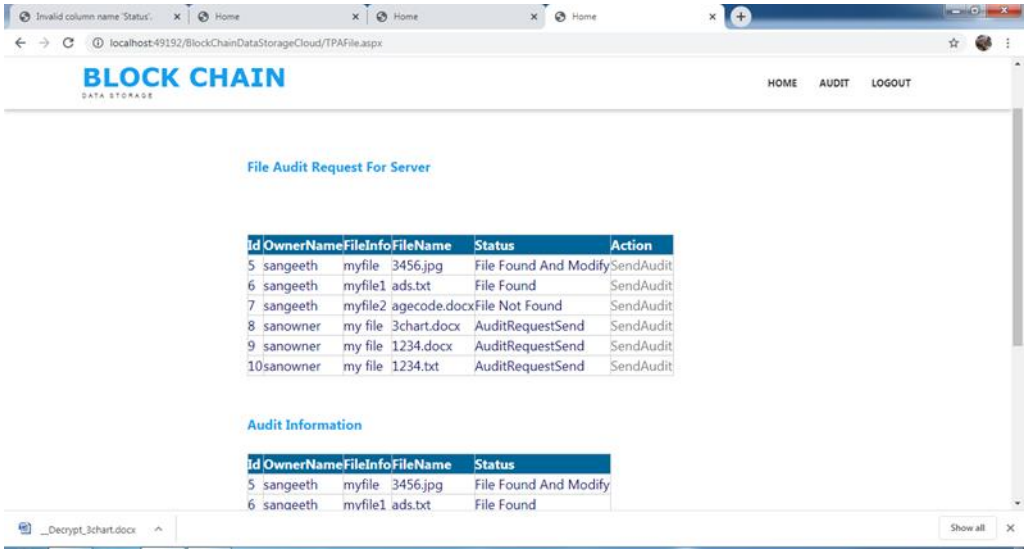

Figure 6. Block request and Verification

Figure 7.shows the file found, file modified or file not found messages. Through this process, the data owner and cloud service providers are easily identify the details of a particular block in an efficient manner.

Figure 7. File found, modified and not found

5. Conclusion

Implemented secure data storage with the help of blockchain based data storage. Provides the strongest security guarantee compared with existing schemes. A procrastinating auditor can detect the data corruption as soon as possible. TPA would not learn any knowledge about the data content stored on the cloud server during the efficient auditing process. Future work focused on TPA may concurrently handle multiple audit sessions from different users for their outsourced data files. Further extend our privacy-preserving public auditing protocol into a multi-user setting, where the TPA can perform multiple auditing tasks in a batch manner for better efficiency.

References

1. Xiaodong Yang, Xixhen Pei, Meiding Wang, Ting Li and Caifen Wang, “Multi-replica and multi-cloud data public audit scheme based on blockchain”, IEEE ACCESS, Volume XX, 2020, PP 1 to 14.
2. Li Yannan, Yong Yu, Bo Yang, Geyong Min, and Huai Wu, “Privacy preserving cloud data auditing with efficient key update”, Future Generation Computer Systems 78(2018), PP 789-798.
3. ShenJian, Jun Shen, Xiaofeng Chen, Xinyi Huang, Willy Susilo, “An efficient public auditing protocol with novel dynamic structure for cloud data”, IEEE Transactions on Information Forensics and Security, 12 No.10, (2017), 2402-2415.
4. ShenWenting, Guangyang Yang, Jia Yu, Hanlin Zhang, Fanyu Kong, RongHao, “Remote data possession checking with privacy preserving authenticators for cloud storage”, Future generation Computer Systems 76, (2017), 136-145.
5. ShenWenting, Jai Yu, Hui Xia, Hanlin Zhang, Xiuqing Lu, RongHao, “Light-weight and privacy preserving secure cloud auditing scheme for group users via the third party medium”, Journal of Network and Computer Applications, Vol.82, (2017), PP 56-64.

6. Yu Yong, Man Ho Au, Geyong Min, "Identity based remote data integrity checking with perfect data privacy preserving for cloud storage", IEEE Transactions on Information Forensics and Security, Vol.12, No.4, (2016), PP 767-778.
7. TianHui, Fulin Nan, Hong Jiang, Chin-Chen Chang, JiantingNing, Yongfeng Huang, "Public auditing for shared cloud data with efficient and secure group management", Information Sciences, No. 472, (2019), PP 107-125.
8. M.Sumathi andS.Sangeetha, "Survey on Sensitive Data Handling- Challenges and Solutions in Cloud Storage System", Advances in Big data and Cloud Computing, PP 1-17, 2019.
9. M.Sumathiand S.Sangeetha, "Scale based secured sensitive data storage for banking services in cloud", International journal of Electronic Business, 14 (2), PP 171-188, Inderscience publisher, 2018.
10. M.Sumathi, S.Sangeetha and Anu Thomas, "Generic cost optimized and secured sensitive attribute storage model for template based text document on cloud", Computer Communication, Vol.150, PP 569-580, Elsevier publisher, 2020.
11. M.Sumathi, R.Lekaa, R.Kavirakshana, N.Nishanthini, K.Nirmala, "Improved CiphertextAttribute Based Sensitive document protection and Secure sharing in cloud storage", International Journal of Advanced Science and Technology, Vol. 29, No.03, PP 8702-8708, 2020.
12. M.Sumathi and S.Sangeetha, "A group key based sensitive attribute protection in cloud storage using modified random Fibonacci cryptography", Complex & Intelligent Systems, PP 1 -15, Springer publisher, 2020.
13. M.Sumathi, S.Sangeetha, "Blockchain based sensitive attribute storage and access monitoring in banking system", International journal of cloud applications and Computing, Vol.10, Article 5, IGI Global Publications, PP 77-92.
14. R.SasiKumar & S.DeviPriya, "Virtual Environment for Treating Anxiety Disorder using GVR Algorithm with Artificial Intelligence", International Journal of Innovative Research in Science Engineering and Technology, pp-2319-8753 Vol.8, Issue 6 June 2019.
15. R.SasiKumar, R.Ramesh, & R.Meena, "Deep learning for Informatics using Fuzzy Logic", International Journal of Innovative Research in Science, Engineering and Technology, pp-2319-8753 Vol.8, Issue.6 June 2019.
16. R.Sasi Kumar & S.DeviPriya,(2019), "Combining Deduplication and String Comparison for Avoiding Redundant Data with Enhanced Authentication Approach on Cloud", International Journal of Advanced Research in Computer and Communication Engineering, Issue6, Vol.8 pp-2278-1021.
17. R.Sasikumar, B.Haritha, T.Borshiya Vincy, M.Kamali & S.De va Priya," Alumni Info-Com with Distinct Classification of Data using Support Vector Machine Algorithm", International Journal of Recent Technology and Engineering (IJRTE), 2277-3878, Volume-8 Issue-6, March 2020.
18. T.M.Nithya, J. Ramya, L. Amudha,"Scope Prediction Utilizing Support Vector Machine for Career Opportunities", International Journal of Engineering and Advanced Technology (IJEAT), ISSN: 2249- 8958, Volume-8 Issue-5, June 2019, pp.2759-2762.
19. L. Amudha, Dr.R.PushpaLakshmi, "Scalable and Reliable Deep Learning Model to Handle Real-Time Streaming Data", International Journal of Engineering and Advanced Technology, ISSN: 2249 – 8958, Volume-9 Issue-3, February, DOI: 10.35940/ijeat.C6272.029320, 2020, Retrieval Number: C6272029320/2020©BEIESP, pp. 3840 – 3844
20. T.M.Nithya, K.S.Guruprakash, L.Amudha. (2020). DEEP LEARNING BASED PREDICTION MODEL FOR COURSE REGISTRATION SYSTEM. International Journal of Advanced Science and Technology, 29(7s), 2178-2184
21. Nithya, T.M., Chitra, S.. (2020). Soft computing-based semi-automated test case selection using gradient-based techniques. Soft Computing. 24. 12981–12987 (2020)
22. K.S.Guruprakash, R.Ramesh, Abinaya K, Libereta A, Lisa Evanjiline L, Madhumitha B. (2020). Optimized Workload Assigning System Using Particle Swarm Optimization. International Journal of Advanced Science and Technology, 29(7), 270
23. T.Vigneshwaran, K.S.Guruprakash, K.Thaslima Nasreen, M.Supraja,(2020), Effective Framework for real time video face recognition system, Journal of Advanced Research in Dynamical and Control system, Vol 12, Issue 6, pp680-684
24. K. S. Guruprakash, M. Jaiganesh, V. Vijey Nathan, R. Sathya. (2021). Location and Temporal based Multi-Cloud Package Selection for Enterprise. Annals of the Romanian Society for Cell Biology, 25(2), 1066 - 1075.