RESEARCH ARTICLE OPEN ACCESS

Different Data Replication Strategies in Cloud Environment

Guilherme Ferreira Gomes
Assistant Mechanical Engineering.Federal Univeristy Of Itajub´A, Itajub´A, Brazil

ABSTRACT
Replication offers high performance, high data availability as well as it enhances the systems reliability. When the data files replica is increased, the maintenance as well as the production cost also goes beyond the limit. More consequently replication technique undergoes on three main limitations. Here, we have performed a comparative study of the different data replication strategies such as ADRS (Adaptive Data Replication Strategy), DCR2S (Dynamic Cost Aware Re-replication and Rebalancing Strategy) and EPA (Efficient Placement Algorithm) in cloud environment. The implementation of these three techniques is done in JAVA and the performance analysis is conducted to study the performance of those replication techniques by using various parameters. The parameters utilized for the comparison of performance of these three techniques are Load variance, Response time, Probability of file availability, SBER (System Byte Effective Rate), Latency, and Fault Ratio. From the analysis, it is evaluated that by varying the number of file replicas it shows deviations on the outcomes of these parameters.
Keywords: Data Replication, Data placement

I. INTRODUCTION

Cloud processing empowers the administration of tremendous IT administrations that is being developed on the topmost of topographically distributed domains and shared it entirely [1-10]. Thus, for better execution and for better reliability, the assets (resources) ought to be replicated over redundant positions as well as it utilizes the redundant frameworks [11-16]. To deploy the data traffics exponential increment and to optimize the datacenter frameworks bandwidth and energy a few data replication techniques is established. Handling the replicas over numerous sites obviously increases the execution thus by diminishing the current access delay and by the elimination of unique failure point [17-26]. In order to maintain the replicas of data, few frameworks, namely the networking devices and storage devices are being demanded. Beyond this, it is necessary to synchronize the currently generated new replicas and any of the progressions created over one of the locales (sites) required to be copied onto another position [27-35].

Data center framework intakes more energy and thus it turns out to be non-utilized. The utilized assets (resources) get to be depleted without including the extra expense. Thus, by the availability of flexible services together with better performance and tremendous data availability turns to be a necessary demand that ought to be met in different situations. In order to face this demand, the term replication is utilized [36]. Data replication is the method of maintaining data files multiple copies over multiple sites thus to enhance better availability, reliability, and performance. Replication is basically utilized in intensive data applications in which the data ought to be shared and required to be gained from distinctive sites, placed over various topographical domains. Replication builds accessibility of the data [37]. If a portion of the sites capturing the information about the file failure at the time of the request, the request will be provided from any other sites handling the replicas (copies) of the specific file [38]. Beyond this, the replication technique also enhances the entire system performance where the request from users is possible to be satisfied over closest sites maintaining the demanded data files replicas. This process maximizes the system's response time as well as the access time. Finally, the replication will also increase the reliability of the system. By chance the corruption of any of the data files replica, or else it fails because of few criteria the request from users can be provided over the different site which is being non-corrupted that maintains the specific replicas [39-46].

II. SYSTEM MODEL

The illustrations in Figure.1 depicts the multi-level hierarchical, heterogeneous cloud framework comprises of various virtual machines, brokers, host, replica catalogs, users and replica managers. By nature the data centers are basically heterogeneous. Various Service Providers controls the data centers. The Service Providers mainly choose about the replication strategies, replication cost and various issues associated with the data centers configuration. Every data center is comprised with different hosts.

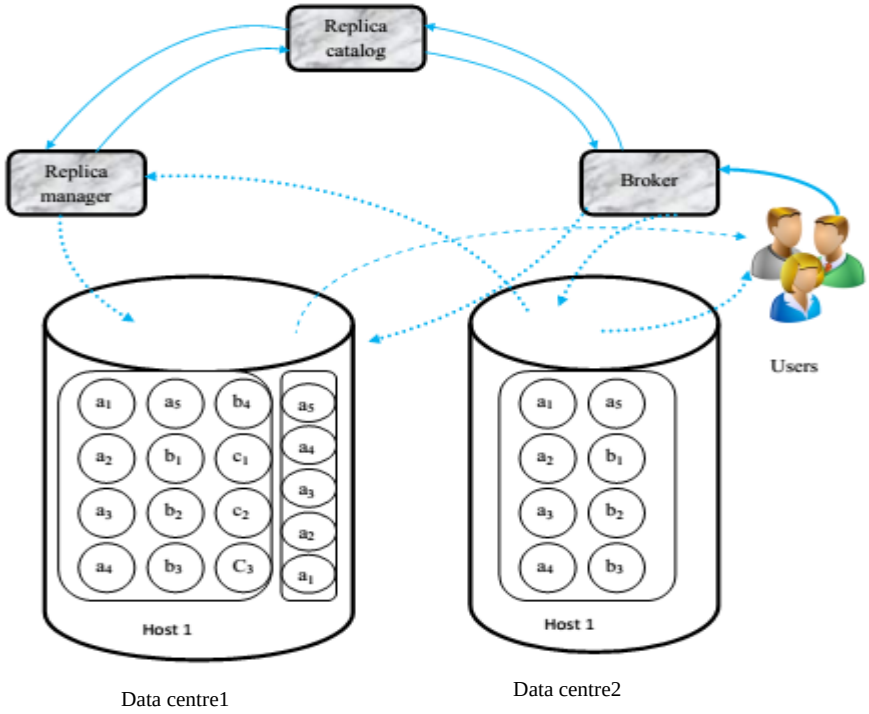

Figure 1: Replication Scenario of Heterogeneous Cloud System

The host is referred to as a physical machine which is being ought to administer various virtual machines. In order to access the data file, the user transmits the demand (request) to the broker. The position of each data files replicas is maintained by replica catalog. Because of this the broker initially transmits the request to the replica catalog in order to access the list of data centers holding the data files that is being requested [47].

After the acceptance of the data centers list capturing the details about the data file that is being requested, the request is then scheduled to the neighbor data center by the broker. Hereby the replica manager is in charge for the generation of new replicas over various domains. The replica manager is additionally in charge of removing the unused as well as old replicas. The block is referred to as the fundamental unit of storage in a data center.

3.1 Dynamic Cost-Aware Re-replication and Re-adjusting Strategy

On cloud frameworks, together with the accessibility, fault maintenance, performance, the replication expense also turns to be a critical component. The method [48] proposed an algorithm referred to as Dynamic Cost-Aware Re-replication and Re-adjusting Strategy (DCR2S) towards a heterogeneous cloud computing framework [49-56]. To begin with these proposed algorithms initially it ought to be designed with heterogeneous cloud framework to comprehend the replication cost, availability of the file, and framework availabilities relationship. The introduced algorithm includes three strides and these strides can be utilized to replicate the data files. At the initial stride, the more popular or efficient data file was selected depending upon the intensity of access together with that the decision has been generated to decide the event when the replication function should be invoked [57-65]. At the second stride, the new replicas ought to be established for the preferred data file thus to face the availability needs.

III. RESULTS AND DISCUSSIONS

Here, we have plotted the graph for all the three techniques with the varying number of replicas in x-axis and the corresponding parameters in y-axis. The performance evaluation is carried out with five different parameters namely the SBER (System Byte Effective Rate), Load, Response time, Latency and Fault Ratio.

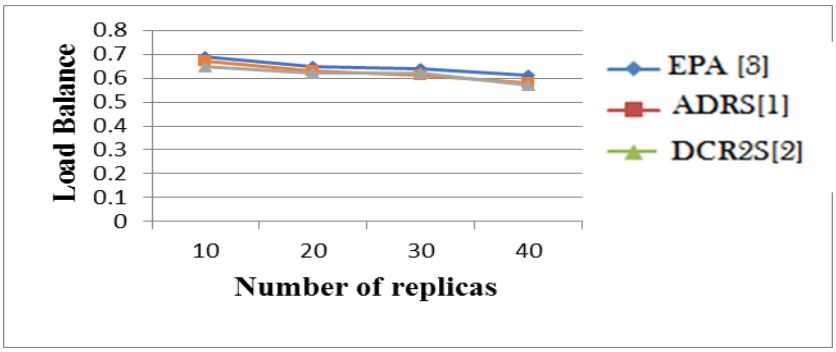

Figure 2: Number of Replicas VS Load

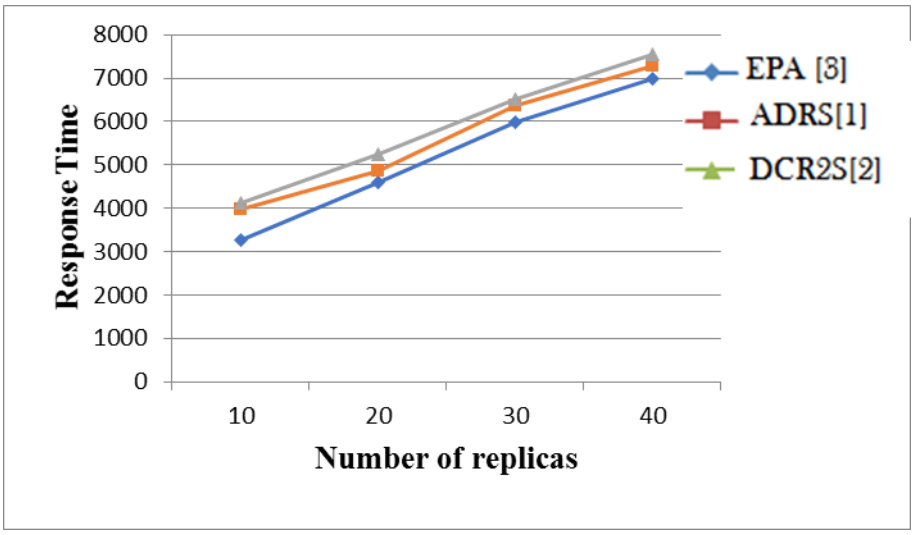

Figure 3: Number of Replicas VS Response Time

Figure.2 and 3 delineates the percentage of load balance when the number of replicas is maximized for about 5 nodes. When number of replicas is increased in the node, the maintenance of load in node turns to be difficult.

IV. CONCLUSION

In this comparative study, we have presented an extensive analysis of three different data replication strategies on a cloud environment. The performance study of these three data replication techniques was carried out in terms five different parameters. Here, the analysis has been made with the help of by varying the number of nodes from 5 to 10 with respect to number of replicas. For all the different analysis, graphs were plotted and the detailed analysis was performed to identify the efficient technique from the three different techniques taken for experimental analysis. Finally, EPA [3] when compared with ADRS [1]and DCR2S [2] provides better performance after evaluating it with different parameters.

REFERENCE

[1] Anand Nayyar, Vikram Puri, Nhu Gia Nguyen, Dac Nhuong Le, Smart Surveillance Robot for the Real Time Monitoring and Control System in Environment and Industrial Applications, Advances in Intelligent System and Computing, pp 229-243, Springer

[2] Ezhilarasu, P., & Krishnaraj, N. (2015). Applications of Finite Automata in Lexical Analysis and as a Ticket Vending Machine–A Review. Int. J. Comput. Sci. Eng. Technol, 6(05), 267-270.

[3] Agrawal, U., Arora, J., Singh, R., Gupta, D., Khanna, A., & Khamparia, A. (2020). Hybrid Wolf-Bat Algorithm for Optimization of Connection Weights in Multi-layer Perceptron. ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM), 16(1s), 1-20.

[4] Prasanna, S., & Ezhilmaran, D. (2016). Association rule mining using enhanced apriori with modified GA for stock prediction. International Journal of Data Mining, Modelling and Management, 8(2), 195-207.

[5] Pustokhina, I. V., Pustokhin, D. A., Gupta, D., Khanna, A., Shankar, K., & Nguyen, G. N. (2020). An effective training scheme for deep neural network in edge computing enabled Internet of medical things (IoMT) systems. IEEE Access, 8, 107112-107123.

[6] Shankar, K., Zhang, Y., Liu, Y., Wu, L., & Chen, C. H. (2020). Hyperparameter tuning deep learning for diabetic retinopathy fundus image classification. IEEE Access, 8, 118164-118173.

[7] Joshi, G. P., Perumal, E., Shankar, K., Tariq, U., Ahmad, T., & Ibrahim, A. (2020). Toward Blockchain-Enabled Privacy-Preserving Data Transmission in Cluster-Based Vehicular Networks. Electronics, 9(9), 1358.

[8] Saračević, M. H., Adamović, S. Z., Miškovic, V. A., Elhoseny, M., Maček, N. D., Selim, M. M., & Shankar, K. (2020). Data Encryption for Internet of Things Applications Based on Catalan Objects and Two Combinatorial Structures. IEEE Transactions on Reliability.

[9] Namasudra, S., & Roy, P. (2017). Time saving protocol for data accessing in cloud computing. IET Communications, 11(10), 1558-1565.

[10] Elsir, A., Elsier, O., Abdurrahman, A., & Mubarakali, A. (2019). Privacy Preservation in Big Data with Data Scalability and Efficiency Using Efficient and Secure Data Balanced Scheduling Algorithm.

[11] Ezhilarasu, P., Krishnaraj, N., & Babu, S. V. (2015). Applications of finite automata in text search-a review. International Journal of Science, Engineering and Computer Technology, 5(5), 116.

[12] Huyen, D.T.T., Binh, N.T., Tuan, T.M., Nguyen, G.N, Dey, N., Son, L.H, Analyzing trends in hospital-cost payments of patients using ARIMA and GIS: Case study at the Hanoi Medical University Hospital, Vietnam, Journal of Medical Imaging and Health Informatics, 7(2), pp. 421-429.

[13] Prasanna, S., & Maran, E. (2015). Stock Market Prediction Using Clustering with Meta-Heuristic Approaches. Gazi University Journal of Science, 28(3).

[14] Pustokhina, I. V., Pustokhin, D. A., Rodrigues, J. J., Gupta, D., Khanna, A., Shankar, K., ... & Joshi, G. P. (2020). Automatic Vehicle License Plate Recognition using Optimal K-Means with Convolutional Neural Network for Intelligent Transportation Systems. IEEE Access.

[15] Namasudra, S. (2018). Cloud computing: A new era. Journal of Fundamental and Applied Sciences, 10(2).

[16] Uthayakumar, J., Elhoseny, M., & Shankar, K. (2020). Highly Reliable and Low-Complexity Image Compression Scheme Using Neighborhood Correlation Sequence Algorithm in WSN. IEEE Transactions on Reliability.

[17] Deepalakshmi, P., & Shankar, K. (2020). Role and Impacts of Ant Colony Optimization in Job Shop Scheduling Problems: A Detailed Analysis. Evolutionary Computation in Scheduling, 11-35.

[18] Ashwin, M., Kamalraj, S., & Azath, M. (2019). Multi objective trust optimization for efficient communication in wireless M learning applications. Cluster Computing, 22(5), 10687-10695.

[19] Ezhilarasu, P., & Krishnaraj, N. (2015). Double Substring based Classification for Nondeterministic Finite Automata. Indian Journal Of Science And Technology, 8, 26.

[20] Amira S. Ashour, Samsad Beagum, Nilanjan Dey, Ahmed S. Ashour, Dimitra Sifaki Pistolla, Gia Nhu Nguyen, Dac-Nhuong Le, Fuqian Shi (2018), Light Microscopy Image De-noising using Optimized LPA-ICI Filter, Neural Computing and Applications, Vol.29(12), pp 1517–1533, Springer, ISSN: 0941-0643.

[21] Prasanna, S., Govinda, K., & Kumaran, U. S. (2012). An Evaluation study of Oral Cancer

Detection using Data Mining Classification Techniques. International Journal of Advanced Research in Computer Science, 3(1).

[22] Sankhwar, S., Gupta, D., Ramya, K. C., Rani, S. S., Shankar, K., & Lakshmanaprabu, S. K. (2020). Improved grey wolf optimization-based feature subset selection with fuzzy neural classifier for financial crisis prediction. Soft Computing, 24(1), 101-110.

[23] Namasudra, S., & Deka, G. C. (2018). Introduction of DNA computing in cryptography. In Advances of DNA computing in cryptography (pp. 1-18). Chapman and Hall/CRC.

[24] Mubarakali, A., Srinivasan, K., Mukhalid, R., Jaganathan, S. C., & Marina, N. (2020). Security challenges in internet of things: Distributed denial of service attack detection using support vector machine-based expert systems. Computational Intelligence.

[25] Le Nguyen Bao, Dac-Nhuong Le, Gia Nhu Nguyen, Vikrant Bhateja, Suresh Chandra Satapathy (2017), Optimizing Feature Selection in Video-based Recognition using Max-Min Ant System for the Online Video Contextual Advertisement User-Oriented System, Journal of Computational Science, Elsevier ISSN: 1877-7503. Vol.21, pp.361-370.

[26] Ezhilarasu, P., Thirunavukkarasu, E., Karuppusami, G., & Krishnaraj, N. (2015). Single substring based classification for nondeterministic finite automata. International Journal on Applications in Information and Communication Engineering, 1(10), 29-31.

[27] Bhateja, V., Gautam, A., Tiwari, A., Nhu, N.G., Le, D.-N, Haralick features-based classification of mammograms using SVM, Advances in Intelligent Systems and Computing, Volume 672, 2018, Pages 787-795.

[28] Latha, A., Prasanna, S., Hemalatha, S., & Sivakumar, B. (2019). A harmonized trust assisted energy efficient data aggregation scheme for distributed sensor networks. Cognitive Systems Research, 56, 14-22.

[29] Krishnaraj, N., Elhoseny, M., Lydia, E. L., Shankar, K., & ALDabbas, O. (2020). An efficient radix trie-based semantic visual indexing model for large-scale image retrieval in cloud environment. Software: Practice and Experience.

[30] Namasudra, S., Roy, P., Vijayakumar, P., Audithan, S., & Balusamy, B. (2017). Time efficient secure DNA based access control model for cloud computing environment. Future Generation Computer Systems, 73, 90-105.

[31] Lakshmanaprabu, S. K., Shankar, K., Rani, S. S., Abdulhay, E., Arunkumar, N., Ramirez, G., & Uthayakumar, J. (2019). An effect of big data technology with ant colony optimization based routing in vehicular ad hoc networks: Towards smart cities. Journal of cleaner production, 217, 584-593.

[32] Namasudra, S., & Deka, G. C. (Eds.). (2018). Advances of DNA computing in cryptography. CRC Press.

[33] Mubarakali, A., Ashwin, M., Mavaluru, D., & Kumar, A. D. (2020). Design an attribute based health record protection algorithm for healthcare services in cloud environment. Multimedia Tools and Applications, 79(5), 3943-3956.

[34] Dey, N., Ashour, A.S., Chakraborty, S., Le, D.-N., Nguyen, G.N, Healthy and unhealthy rat hippocampus cells classification: A neural based automated system for Alzheimer disease classification, Journal of Advanced Microscopy Research, 11(1), pp. 1-10

[35] Krishnaraj, N., Ezhilarasu, P., & Gao, X. Z. Hybrid Soft Computing Approach for Prediction of Cancer in Colon Using Microarray Gene Data. Current Signal Transduction Therapy, 11(2).

[36] Namasudra, S., Deka, G. C., Johri, P., Hosseinpour, M., & Gandomi, A. H. (2020). The revolution of blockchain: State-of-the-art and research challenges. Archives of Computational Methods in Engineering.

[37] Goel, N., Grover, B., Gupta, D., Khanna, A., & Sharma, M. (2020). Modified Grasshopper Optimization Algorithm for detection of Autism Spectrum Disorder. Physical Communication, 101115.

[38] Prasanna, S., Narayan, S., NallaKaruppan, M. K., Anilkumar, C., & Ramasubbareddy, S. (2019). Iterative Approach for Frequent Set Mining Using Hadoop Over Cloud Environment. In Smart Intelligent Computing and Applications (pp. 399-405). Springer, Singapore.

[39] Le, D.-N.a, Kumar, R.b, Nguyen, G.N., Chatterjee, J.M.d, Cloud Computing and

Virtualization, DOI: 10.1002/9781119488149, Wiley.

[40] Raj, R. J. S., Shobana, S. J., Pustokhina, I. V., Pustokhin, D. A., Gupta, D., & Shankar, K. (2020). Optimal Feature Selection-Based Medical Image Classification Using Deep Learning Model in Internet of Medical Things. IEEE Access, 8, 58006-58017.

[41] Namasudra, S., & Deka, G. C. (2018). Taxonomy of DNA-based security models. In Advances of DNA Computing in Cryptography (pp. 37-52). Chapman and Hall/CRC.

[42] Mubarakali, A., Ramakrishnan, J., Mavaluru, D., Elsir, A., Elsier, O., & Wakil, K. (2019). A new efficient design for random access memory based on quantum dot cellular automata nanotechnology. Nano Communication Networks, 21, 100252.

[43] Ramakrishnan, J., Mavaluru, D., Sakthivel, R. S., Alqahtani, A. S., Mubarakali, A., & Retnadhas, M. (2020). Brain–computer interface for amyotrophic lateral sclerosis patients using deep learning network. NEURAL COMPUTING & APPLICATIONS.

[44] Van, V.N., Chi, L.M., Long, N.Q., Nguyen, G.N., Le, D.-N, A performance analysis of openstack open-source solution for IaaS cloud computing, Advances in Intelligent Systems and Computing, 380, pp. 141-150.

[45] Sinha, A., Shrivastava, G., Kumar, P., & Gupta, D. (2020). A community-based hierarchical user authentication scheme for Industry 4.0. Software: Practice and Experience.

[46] Namasudra, S., Devi, D., Kadry, S., Sundarasekar, R., & Shanthini, A. (2020). Towards DNA based data security in the cloud computing environment. Computer Communications, 151, 539-547.

[47] Mubarakali, A., Durai, A. D., Alshehri, M., AlFarraj, O., Ramakrishnan, J., & Mavaluru, D. (2020). Fog-Based Delay-Sensitive Data Transmission Algorithm for Data Forwarding and Storage in Cloud Environment for Multimedia Applications. Big Data.

[48] Reshmi, T. R., & Azath, M. (2020). Improved self-healing technique for 5G networks using predictive analysis. Peer-to-Peer Networking and Applications, 1-17.

[49] Namasudra, S., Chakraborty, R., Majumder, A., & Moparthi, N. R. (2020). Securing multimedia by using DNA based encryption in the cloud computing environment. ACM Transactions on Multimedia Computing Communications and Applications.

[50] Patro, K. K., Reddi, S. P. R., Khalelulla, S. E., Kumar, P. R., & Shankar, K. (2020). ECG data optimization for biometric human recognition using statistical distributed machine learning algorithm. The Journal of Supercomputing, 76(2), 858-875.

[51] Rajagopal, A., Joshi, G. P., Ramachandran, A., Subhalakshmi, R. T., Khari, M., Jha, S., ... & You, J. (2020). A Deep Learning Model Based on Multi-Objective Particle Swarm Optimization for Scene Classification in Unmanned Aerial Vehicles. IEEE Access, 8, 135383-135393.

[52] Chakchai So-In, Tri Gia Nguyen, Gia Nhu Nguyen: Barrier Coverage Deployment Algorithms for Mobile Sensor Networks. Journal of Internet Technology 12/2017; 18(7):1689-1699.

[53] Mubarakali, A., Bose, S. C., Srinivasan, K., Elsir, A., & Elsier, O. (2019). Design a secure and efficient health record transaction utilizing block chain (SEHRTB) algorithm for health record transaction in block chain. Journal of Ambient Intelligence and Humanized Computing, 1-9.

[54] Devaraj, A. F. S., Murugaboopathi, G., Elhoseny, M., Shankar, K., Min, K., Moon, H., & Joshi, G. P. (2020). An Efficient Framework for Secure Image Archival and Retrieval System Using Multiple Secret Share Creation Scheme. IEEE Access, 8, 144310-144320.

[55] Mubarakali, A. (2020). Healthcare Services Monitoring in Cloud Using Secure and Robust Healthcare-Based BLOCKCHAIN (SRHB) Approach. MOBILE NETWORKS & APPLICATIONS.

[56] Namasudra, S. (2019). An improved attribute-based encryption technique towards the data security in cloud computing. Concurrency and Computation: Practice and Experience, 31(3), e4364.

[57] Kathiresan, S., Sait, A. R. W., Gupta, D., Lakshmanaprabu, S. K., Khanna, A., & Pandey, H. M. (2020). Automated detection and classification of fundus diabetic retinopathy images using synergic deep learning model. Pattern Recognition Letters.

[58] Govinda, K., & Prasanna, S. (2015, February). Medical dialysis prediction using fuzzy rules.

In 2015 International Conference on Soft-Computing and Networks Security (ICSNS) (pp. 1-5). IEEE.

[59] Sujatha, R., Navaneethan, C., Kaluri, R., & Prasanna, S. (2020). Optimized Digital Transformation in Government Services with Blockchain. In Blockchain Technology and Applications (pp. 79-100). Auerbach Publications.

[60] Govinda, K., & Prasanna, S. (2015, February). A generic image cryptography based on Rubik's cube. In 2015 International Conference on Soft-Computing and Networks Security (ICSNS) (pp. 1-4). IEEE.

[61] Anand Nayyar, Vikram Puri, Nhu Gia Nguyen, BioSenHealth 1.0: A Novel Internet of Medical Things (IoMT) Based Patient Health Monitoring System, Lecture Notes in Networks and Systems. Springer, 2019

[62] Prasanna, S., & Narayanan, V. (2017). A Novel Approach for Generation of All-Optical OFDM Using Discrete Cosine Transform Based on Optical Couplers in a Radio-Over-Fiber Link. International Journal of Advanced Research in Engineering and Technology, 8(3).

[63] Sanjeevi, P., Prasanna, S., Siva Kumar, B., Gunasekaran, G., Alagiri, I., & Vijay Anand, R. (2020). Precision agriculture and farming using Internet of Things based on wireless sensor network. Transactions on Emerging Telecommunications Technologies, e3978.

[64] Rathi, V. K., Chaudhary, V., Rajput, N. K., Ahuja, B., Jaiswal, A. K., Gupta, D., ... & Hammoudeh, M. (2020). A Blockchain-Enabled Multi Domain Edge Computing Orchestrator. IEEE Internet of Things Magazine, 3(2), 30-36.

[65] Khanna, A., Rodrigues, J. J., Gupta, N., Swaroop, A., & Gupta, D. (2020). Local mutual exclusion algorithm using fuzzy logic for Flying Ad hoc Networks. Computer Communications.

[66] Satish, Karuturi S R V, and M Swamy Das. "Multi-Tier Authentication Scheme to Enhance Security in Cloud Computing." IJRAR (International Journal of Research and Analytical Reviews) 6, no. 2 (2019): 1-8, 2019.