An Analysis of Privacy Preserving Data Storage and Retrieval Approaches in Heterogeneous Multi-cloud Architectures

Suganya M¹

¹Research Scholar, Department of Computer Science and Engineering, Sathyabama Institute of Science and Technology, Chennai, Tamil Nadu, India

¹Assistant Professor, Department of Computer Science and Engineering, Jeppiaar Institute of Technology suganyam02@gmail.com

Dr.T.Sasipraba²

²Vice Chancellor, Sathyabama Institute of Science and Technology, Chennai, Tamil Nadu, India

vc@sathyabama.ac.in

ABSTRACT: Cloud storage and retrieval are considered essential services of cloud computing (CC), which enables the data owner to store the data from the local computing systems to the cloud. Though cloud storage provides several benefits such as minimal cost and short turnaround time, data security is found to be a highly challenging and difficult process. One of the effective solutions to avoid data loss related to security risks is the concurrent usage of multi-cloud. The multi-cloud allows the data owner to easily access their data from remote areas via the web interfaces offered by Amazon EC2. However, even if such multi-cloud environments are susceptible to security attacks, it would bring irreversible challenges to the owners namely integrity, availability, and confidentiality of data. In order to resolve these issues, several privacy preserving data storage and retrieval models are presented in the literature to store data securely and reduce the computational resources. In this view, this study performs a review of existing state of art data storage and retrieval approaches for heterogeneous multi-cloud architectures. The reviewed methods are investigated based on their objectives, underlying techniques, implementation data, and evaluation parameters. Besides, a comparative results analysis of few of the reviewed models takes place. Finally, a discussion along with the possible future direction is given to enable the readers to find the research problem.

Keywords: Cloud service providers, Data retrieval, Data Storage, Multi-cloud, Privacy preserving, Security

1. INTRODUCTION

Cloud computing (CC) is most developing technology that offers a lot of advantages for the organizations and common people to store and share files. Scalability, flexibility, global access, and availability are vital to these files. The CC module decreases the maintenance cost and infrastructure by offering secure services and computing at lower cost [1, 2]. Although CC provides several benefits, service availability failure, security and privacy are the crucial problems that should be tackled. Based on the current study it moved to a multi cloud [3]. The Multi cloud is the computing module which combines several computing services to a single heterogeneous framework for minimizing the threat of data loss, service availability failure, loss of privacy, vendor lock-in corruption of data, and chance of malicious insiders from single cloud. In the traditional data storage module, an organization should acquire communication devices, servers, storage components, maintain the centre and human resources to construct. In contrast, multi cloud storage provides clients additional comfortable, flexibility, and scalability on data storage at lower rates, depending upon the needs of memory instead of knowing how to create and maintain the storage. An instance of public cloud storage service given by data acquisition and various organizations [4]. Fig. 1 shows the architecture of multicloud system.

Multi-Cloud is the use of many computing services in a single heterogeneous framework. The Multi-Cloud Storage implies the usage of many cloud storage services by a single web interface instead of defaults given by cloud storage vendor in a single heterogeneous framework. A Multi-Cloud data system can improve data sharing and this feature would be considerably assisted the user information. It allows data owners for sharing its information from cloud. Over the entire CC model, security is considered as one of the critical features

because of the delicacy and sensitivity of the user data or information stored in a cloud. At present, all Organizations are moving its IT sector for scaling up its data sharing system.

Most of the cloud services aren't open and possess distinct sizes. For example, Single Cloud Storage falls between the services with storage constraint that creates it harmful compared with multi cloud save. Businesses must begin to consider functioning with several cloud providers during performance, disaster recovery, cost savings, etc. Most business organizations share their information with their suppliers/clients and deliberate data sharing as a priority [5]. With many clients from several organizations contribute to the cloud data, cost and time spent will be lesser than conventional way of automatically sharing and sending information that frequently led to the formation of redundant and outdated documents.

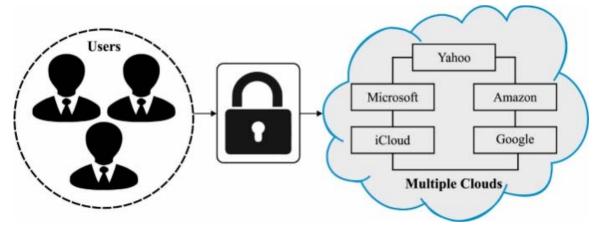


Fig 1 Data Storage in Multi Cloud Architecture

Though various cryptographic data slicing approaches [6] were presented as the major challenge appear in the insider access for storing information. Insiders are the confidential second manager/admin which preserves the third-party server with similar authority as the admin. As the third-party infrastructure/servers are utilized for storing various sensitive information's. Administrator and third party manages the organization as they have remote access to the server; when administrator/third parties manager is malicious then they attain access for the data users. Another risk is different from single cloud storage, recovery of the sliced files from multi cloud server isn't an easier process.

Additionally, malicious files could be easily download in any present method in single and multi-cloud storage [7]. The reduced emphasis was employed in implementing the multi cloud framework if malicious files are downloaded. Present, reliability assurances only based on every each CSP's individual backup resolution can be a single point of failure. Most critically, this crash has enduringly damaged several user's data with severe consequences to few clients [8, 9]. It is noteworthy that a complete resolution for ensuring simultaneous privacy, reliability, and search ability on information deployed to many clouds isn't unimportant for determining. Simple data replicating at several CSPs is one of the direct ways that is the minimum cost efficient method.

2. MOTIVATION

According to recent research, the amount of data in the world is 44 zettabytes. Organizations and government sectors prefer to store data in cloud rather than in local machine. Securing data in the cloud is very important aspect to be considered as intruders and attackers get access to the sensitive data during storage or retrieval at the end user site. In a statistical study, about 29% is affected by data breaches affecting overall 93% of the organizations. To overcome vulnerabilities and safeguard the data from breaches and to reduce threats globally, efficient privacy preserving data storage and retrieval approaches need to be implemented.

3.REVIEW OF EXISTING SECURITY RELATED SOLUTIONS FOR MULTI-CLOUD ARCHITECTURES

This section investigates the different security based solutions related to data storage and retrieval in multicloud. Lahmar and Mezni [10] developed a security aware multi cloud service composition method by RS and fuzzy formal concept analyses (fuzzy FCA), that are 2 methods with a robust mathematic background. To assurance a higher security level of the hosting cloud and the chosen service, they utilize FCA and calculation of RS. This method would assist decreasing the search space, with the elimination of insecure services and disqualified clouds. The investigational outcomes showed the effectiveness and performance of this method.

In Pachala et al. [11], a hybrid method with multi cloud hosting platform is implemented and designed to improve privacy and security of cloud data. This technique contains (i) Byzantine protocol for tolerating security breaches for server failure cloud that is autonomous. (ii) DepSky framework improves the secrecy and reliability of data maintained in the cloud by decoding and encoding methods (iii) Shamir secret sharing process to enhance privacy and trustiness of data storage without influencing the efficiency. The security and privacy problems of hybrid method are executed and related to the protocols such as SAML with Kerberos and proxy reencryption for various client services request. Torkura et al. [12] proposed CSBAuditor, a new cloud security scheme which always monitor cloud framework, to identify malicious activity and unauthorized modifications. The CSB Auditor leverages 2 ideas: state transition analyses and reconciler pattern for overcoming the above mentioned security problems. Moreover, security metrics are utilized for computing severity score to detect vulnerability by a new scoring scheme: Cloud Security Scoring Scheme.

Zhu et al. [13] proposed a new scheduling technique named MMA for optimizing make span and overall costs for entire submitted process subject to reliability and security limitations. This technique is separated to 2 stages for scheduling task. The initial stage is to detect an optimum matching candidate resource to the task for meeting their superior requirements like security, reliability and performance in a multi cloud platform; next it iteratively executes many rounds of reallocating for optimizing task performance time and cost by reducing the difference of the calculated finishing time. The presented method, modified artificial bee colony (MABC), hybrid chaotic particle search (HCPS), max-min, and min-min, modified cuckoo search (MCS), algorithms are executed in CloudSim for creating simulation.

Megouache et al. [14] presented a novel module for solving security problems in this platforms. This method contains of 3 phases, the initial stage, is for proposing a private virtual network to secure the data transmission. Next, they utilized an authentication technique depending upon data encryption, for protecting the user identity and his information, and lastly, they understand a method for knowing the reliability of data allocated on the several clouds of the scheme. The module attains identity verification and capability to interoperate among processes run on distinct cloud providers. A data integrity method would be established.

Viswanath and Krishna [15] aim is to improve the secure architecture that restrict the insider attacks. The presented architecture comprises encryption, data uploading, indexing, slicing, decryption, distribution, merging and retrieval procedure. The hybrid encryption method was established for providing the privacy to the big data earlier stored it into the multi cloud. The research analyses is performed by real world cloud storage platform. The presented method record approximately 2630 KB/S for processing encryption.

Cao et al. [16] implement and design a multi cloud architecture to build Open Stack based environment for medicinal IoT, denoted by Tri-SFRS. For implementing this technique, they integrate various methods for attaining this decrease in efforts, comprising lower overhead native testing architecture, multi cloud cascading framework, snapshot volume cascaded operation for b-ultrasonic data and medicinal data storage backup method. Tri-SFRS can concurrently allow assets managing. Tri-SFRS was implemented as native element in the Open Stack environment, and it determines the degree of native Open Stack multi cloud environment management using this presented cascading architecture.

Celesti et al. [17] deliberate to improve the whole systems regarding retrieval and data storage via validating and testing a MCS scheme consist of 3 main Cloud Storage suppliers; Copy, Dropbox and Google Drive. Research have showed that the select of Cloud storage provider for storing files according to data transfer efficiency depends on file chunk size. Pravin et al. [18] to improve the security and privacy of the data in the multi cloud, a strong method is proposed. The major emphasis of this research is to tackle the security and privacy threats of data in the multi cloud storage. Particularly, data is split to several slices. An amount of slices is determined with data owner. The sliced files are encrypted with elliptical curve cryptography (ECC) and three DES (data encryption standard) method. The efficiency of the presented method was estimated by latency time.

Rios et al. [19] proposed a new DevOps architecture intended at assisting Cloud consumer in deploying, designing and functioning (multi) Cloud systems which contain the required security and privacy controls to ensure law enforcement authorities, transparency for end users and third-party in service provisions. The architecture is based on the risk driven requirement at implementation time of security and privacy levels objective in the continuous enforcement and service level agreement and observing at run-time.

Torkura et al. [20] presented a 2 pronged method for automatic threat detection and incident response in multi cloud storage system. The initial method includes dynamic recovery and snapshotting approaches for detecting and partly neutralize security event. The next method build in the first stage with automatic relating the created alert by cloud event log, for extracting actionable intelligence to incident response. Therefore, malicious activity is eliminated, investigated and identified. This method is designed in SlingShot, extend this early study –

CSBAuditor that implement the initial step. The developed methods collaborate in real-time for mitigating above mentioned security problems on Google Cloud Platform (GCP) and Amazon Web Services (AWS).

Casola et al. [21] proposed a new security driven method for the deployment, design and development of multi cloud applications. It can be depending upon fully automatable procedure which helps the developer in elicitation of application needs up for detection of an optimum deployment configuration, allows to detect an optimum compromise among entire costs and attained level of security. The developed optimization procedure takes explicitly into account 2 crucial features that are frequently ignored in related methods, such as cloud ondemand leasing module to resource allocation and the effect that deployment on security strategies are executed by a difficult application.

Tchernykh et al. [22] presented a multi cloud based storage framework named WA-RRNS which integrates threshold secret allocation redundant and weight access system remains number scheme with many failure recognition or recovery mechanism and homomorphic cipher. For an optimum trade-offs among security and efficiency, WA-RRNS utilizes variables for adjusting data loss probability, redundancy and encryption-decryption speed. Investigational and Theoretical analyses with actual data displays that this method gives a secure manner for mitigating the uncertainty of untrusted and not consistent cloud storage.

References	Year	Objective	Technique used	Dataset	Evaluation Metrics	
Lahmar and Mezni [10]	2021	Achieve security in multicloud systems	Fuzzy FCA, RS	-	-	
Pachala et al. [11]	2021	To achieve security and privacy in cloud data	Hybrid technique	-	Memory consumption, encryption/decryption time, total authentication on time	
Torkura et al. [12]	2021	To distinguish malicious activity and illegal change	CSBAuditor		Detection rate, response time, latency	
Zhu et al. [13]	2021	Design a scheduling scheme to optimize makespan and total cost	MCS, HCPS, MABC		Make span, cost and resource utilization	
Megouache et al. [14]	2020	To achieve data confidentiality and integrity in mutlicloud	VPN, RSA	Public data and confidential data of insured person	Download time, Processor usage	
Viswanath and Krishna [15]	2020	Secure model to restrict the insider attacks	Hybrid encryption technique	Real time health data from web site	Throughput, running time, encryption and decryption time	
Cao et al. [16]	2019	Design tri-storage failure recovery system	Tri-SFRS	Medical IoT data	Latency, overhead time	
Celesti et al. [17]	2019	Optimizing the storage and retrival efficiency in multicloud	HMSC module	Real testbed from Google Drive, Dropbox, and Copy	Mean upload and download time	
Pravin et al. [18]	2019	Achieve security in multicloud systems	Dynamic file slicing, 3DES, ECC	Different file types	File Uploading Latency Time (FULT) and File Downloading Latency Time (FDLT)	

Rios et al. [19]	2019	SLA based security and privacy scheme in cloud and multicloud	DevOps approach	Flight Scheduling application,T ampere Smart mobility (TSM) application	-	
Torkura et al. [20]	2019	Detect threats and incident responses automatically	SlingShot	Attack scenario using CloudGoat	Mean time	
Casola et al. [21]	2018	Achieve security in multicloud systems	security-by-design Approach	Simple cloud application, Chat service	-	
Tchernykh et al. [22]	2018	To store data securely in multicloud systems	WA-RRNS	Data from IaaS public cloud, 2015	Probability of denial of access, Encoding and decoding speed	
Subramanian , and Leo [23]	2017	Reduce malicious insider and file threat in multicloud systems	SDSMC	You Tube Dataset	Slicing time, Encryption time, Decryption time, Merge time	
Li et al. [24]	2015	Achieve privacy in multicloud systems	STRE	Enron email dataset	Encryption time, Transmission time	

Table 1 Comparative study of different security based solutions for multicloud environment

Subramanian and Leo [23] goals at providing a framework that decreases malicious insiders and file risks that enhances data sharing security in Multi Cloud storage services. This method would provide a secured platform where the data owner could retrieve and store data from Multi Cloud platform with no merging file conflict and prevent insiders attack for obtaining useful data. Research indicates that the recommended module is appropriate for making decision procedure to the data owners in an optimum acceptance of multi cloud storage service to share their data safely. Li et al. [24] proposed a privacy preserving STorage and REtrieval (STRE) method that guarantees privacy and security however also offers consistency assurances for the outsourced searchable encrypted data. The STRE method allows the cloud user for distributing and searching its encryption data over many independent clouds handled by distinct CSPs, and strong while a specific amount of CSP crashes. In addition to reliability, STRE provides the advantage of partly hidden search pattern. A brief comparison of the reviewed models is given in Table 1.

4.DISCUSSION AND FUTURE DIRECTIONS

In this section, the security analysis of different privacy preserving techniques takes place in Table 2 and Fig. 2 [18, 23]. From the obtained results, it is obvious that the SSDSMC, DSMC, and CP-ABE methods offered higher privacy over the MCPCTA-ABE and SeDASC. In addition, the CP-ABE and SeDASC have identified the insider attacks effectively over the other methods in a considerable way.

Features	MCPCTA- ABE	SSDSMC	DSMC	CP- ABE	SeDaSC
Privacy	40	60	60	60	40
Insider attacks	40	30	60	80	80
Confidentiality	70	70	80	30	30
Secret keys	0	0	0	60	60
Data integrity	20	20	20	20	20

Table 2 Security Analysis (%) of Various Methods

Followed by, the DSMC technique has accomplished maximum confidentially over the other methods. In line with this, the CP-ABE and SeDaSC techniques have offered enhanced performance in the detection of secret keys. Finally, all the compared methods have demonstrated equivalent performance in terms of data integrity.

For instance, the SSDSMC, DSMC, and CP-ABE methods have obtained higher privacy of 60% whereas the MCPCTA-ABE and SeDASC methods have showcased lower privacy of 40%. Besides, the insider attacks are detected by the CP-ABE and SeDASC techniques with the maximum security of 80% whereas the MCPCTA-ABE, SSDSMC, and DSMC techniques have achieved reduced security of 40%, 30%, and 60% respectively. Moreover, the DSMC technique has resulted in maximum confidentiality of 80% whereas the MCPCTA-ABE, SSDSMC, CP-ABE, and SeDaSC techniques have showcased reduced confidentiality of 70%, 70%, 30%, and 30% respectively. Furthermore, the CP-ABE and SeDASC techniques have recognized the secret keys with 60%. Lastly, all the existing methods such as MCPCTA-ABE, SSDSMC, DSMC, CP-ABE, and SeDaSC techniques have offered identical data integrity of 20%.

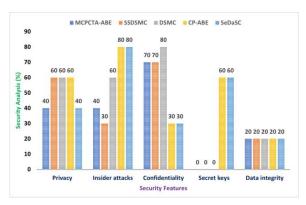


Fig 2 Comparison Analysis Graph

Some of the possible future directions in the multicloud architectures are discussed here. The privacy preserving data storage in multicloud can be realized in several real time applications such as healthcare, education, etc. For instance, the university authorities can share secret data with its affiliated colleges such as question papers. Similarly, business people can utilize it to transfer secret data with clients. In addition, new methods are needed to support virtualization and guarantee bandwidth for multi-tenant datacenter networks. It is also needed to explore more automatic service provisioning techniques guaranteeing Qos. Additional studies are required in the design of load balancing techniques, energy-efficient resource management, and resource scheduling techniques in multicloud environment.

5. CONCLUSION AND FUTURE SCOPE

In recent years, multicloud become a familiar topic and several privacy preserving data storage and retrieval models are presented in the literature to achieve security and privacy. This paper offered a comprehensive review of available data storage and retrieval approaches for heterogeneous multicloud architectures. Each and every reviewed method is examined depending upon the aim, underlying techniques, implementation data, and evaluation parameters. In addition, a detailed comparison study is made with some of the surveyed techniques interms of different measures.

An ability to switch between different cloud providers, such as multiple clouds or federated clouds, could help solve these problems by providing users with alternatives if there is scheduled maintenance, a breach, or a shutdown. Multi-clouds and federated clouds each have their own advantages and disadvantages. One of the great things about hybrid systems is that they are easy to modify to a certain application, but are less transferrable. This makes them less useful in some cases. For enterprises that need numerous jobs or services, multi-cloud and federated clouds are more suitable. The detailed comparison study is made with some of the surveyed techniques in terms of different measures and Finally Future work should embrace multi-cloud paradigms and employ these other technologies, such as machine learning and big data, for new techniques of analysis and a secure framework for data storage and retrieval of data in multi cloud environment can be implemented.

REFERENCES

- [1] Subashini, S. and Kavitha, V., 2011. A survey on security issues in service delivery models of cloud computing. Journal of network and computer applications, 34(1), pp.1-11.
- [2] Li, J., Zhang, Y., Chen, X. and Xiang, Y., 2018. Secure attribute-based data sharing for resource-limited users in cloud computing. Computers & Security, 72, pp.1-12.
- [3] Neelakandan, S. and Muthukumaran, S., Transformation-based Optimizations Framework (ToF) for Workflows and its Security issues in the Cloud Computing.
- [4] Neelakandan, S., Paulraj, D. and Dineshkumar, M., 2015. Decentralized Access Control Of Data In Cloud Services Using Key Policy Attribute Based Encryption, International Journal for Scientific Research & Development, ISSN 2321 – 0613& 3(2).

- [5] Thilakanathan, D., Chen, S., Nepal, S. and Calvo, R.A., 2014. Secure data sharing in the cloud. In Security, privacy and trust in cloud systems (pp. 45-72). Springer, Berlin, Heidelberg.
- [6] Fabian, B., Ermakova, T. and Junghanns, P., 2015. Collaborative and secure sharing of healthcare data in multi-clouds. Information Systems, 48, pp.132-150.
- [7] Neelakandan, S. and Paulraj, D., 2020. An automated exploring and learning model for data prediction using balanced CA-SVM. Journal of Ambient Intelligence and Humanized Computing, pp.1-12.
- [8] Blodget, H., 2011. Amazon's cloud crash disaster permanently destroyed many customers' data. Business Insider.
- [9] Uthayakumar, J., Elhoseny, M. and Shankar, K., 2020. Highly reliable and low-complexity image compression scheme using neighborhood correlation sequence algorithm in WSN. IEEE Transactions on Reliability, 69(4), pp.1398-1423.
- [10] Lahmar, F. and Mezni, H., 2021. Security-aware multi-cloud service composition by exploiting rough sets and fuzzy FCA. Soft Computing, 25(7), pp.5173-5197.
- [11] Pachala, S., Rupa, C. and Sumalatha, L., 2021. An improved security and privacy management system for data in multi-cloud environments using a hybrid approach. Evolutionary Intelligence, pp.1-17.
- [12] Torkura, K.A., Sukmana, M.I., Cheng, F. and Meinel, C., 2021. Continuous auditing and threat detection in multi-cloud infrastructure. Computers & Security, 102, p.102124.
- [13] Zhu, Q.H., Tang, H., Huang, J.J. and Hou, Y., 2021. Task Scheduling for Multi-Cloud Computing Subject to Security and Reliability Constraints. IEEE/CAA Journal of AutomaticaSinica, 8(4), pp.848-865.
- [14] Megouache, L., Zitouni, A. and Djoudi, M., 2020. Ensuring user authentication and data integrity in multi-cloud environment. Human-centric Computing and Information Sciences, 10, pp.1-20.
- [15] Viswanath, G. and Krishna, P.V., 2020. Hybrid encryption framework for securing big data storage in multi-cloud environment. Evolutionary Intelligence, pp.1-8.
- [16] Cao, R., Tang, Z., Liu, C. and Veeravalli, B., 2019. A scalable multicloud storage architecture for cloud-supported medical internet of things. IEEE Internet of Things Journal, 7(3), pp.1641-1654.
- [17] Celesti, A., Galletta, A., Fazio, M. and Villari, M., 2019. Towards hybrid multi-cloud storage systems: Understanding how to perform data transfer. Big Data Research, 16, pp.1-17.
- [18] Pravin, A., Jacob, T.P. and Nagarajan, G., 2019. Robust technique for data security in multicloud storage using dynamic slicing with hybrid cryptographic technique. Journal of Ambient Intelligence and Humanized Computing, pp.1-8.
- [19] Rios, E., Iturbe, E., Larrucea, X., Rak, M., Mallouli, W., Dominiak, J., Muntés, V., Matthews, P. and Gonzalez, L., 2019. Service level agreement-based GDPR compliance and security assurance in (multi) cloud-based systems. IET Software, 13(3), pp.213-222.
- [20] Torkura, K.A., Sukmana, M.I., Cheng, F. and Meinel, C., 2019, September. Slingshot-automated threat detection and incident response in multi cloud storage systems. In 2019 IEEE 18th International Symposium on Network Computing and Applications (NCA) (pp. 1-5), IEEE.
- [21] Casola, V., De Benedictis, A., Rak, M. and Villano, U., 2018. Security-by-design in multi-cloud applications: An optimization approach. Information Sciences, 454, pp.344-362.
- [22] Tchernykh, A., Babenko, M., Miranda-López, V., Drozdov, A.Y. and Avetisyan, A., 2018, May. WA-RRNS: Reliable data storage system based on multi-cloud. In 2018 IEEE International Parallel and Distributed Processing Symposium Workshops (IPDPSW) (pp. 666-673). IEEE.
- [23] Subramanian, K. and Leo, J., 2017. Enhanced Security for Data Sharing in Multi Cloud Storage (SDSMC). Int. J. Adv. Comput. Sci. Appl, 8, pp.176-185.
- [24] Li, J., Lin, D., Squicciarini, A.C., Li, J. and Jia, C., 2015. Towards privacy-preserving storage and retrieval in multiple clouds. IEEE Transactions on Cloud Computing, 5(3), pp.499-509.

Authors Profile



Suganya M, Research Scholar, Department of Computer Science and Engineering,inSathyabama Institute of Science and Technology and working as Assistant Professor, Department of Computer Science and Engineering in Jeppiaar Institute of Technology. She has 5+ years of teaching and Industry Experience. Her Area of Interest is Cloud Security and Satellite Technology. She has been awarded "Young Research Engineer" for carrying out vibrant activities on small satellites under the banner of UNISEC India. She is one of the Project Head for UNITY SAT JIT-SAT launched from ISRO on Feb 28th 2021. She has also been awarded "Excellence in research and Innovation of the Year 2021.



Dr.T. Sasipraba, obtained her B.E and M.E., from the University of Madras and Ph.D from Sathyabama University. She joined Sathyabama University in 1995 as a Lecturer and her 19 years of meritorious career in the same University has promoted her as Vice Chancellor of the university in the year 2020. During the course of her career at Sathyabama University Dr. T. Sasipraba has made exceptional contributions in the areas of research and developments, international linkages and Publications. For her outstanding contributions over the years, Dr. Sasipraba has received numerous awards from Sathyabama University and from Cognizant Technology Solutions. She has published more than 125 papers in refereed international journals and conference proceedings and has guided many Ph.D Scholars in the field of Computer Science and Engineering.