CONSTRUCTION OF CLOUD BASED E-LEARNING PLATFORM WITH PRIVACY PRESERVED PUBLIC DATA VERIFICATION SERVICES

Ms. S. Vinitha¹ and Dr. G. Kesavaraj²

¹M.phil Full Time Research Scholar, PG & Research Department of Computer Science, Vivekanandha College of Arts and Sciences for Women.(Autonomous), Tiruchengode, India. Namakkal 637205.
²Head of the Department, PG & Research Department of Computer Science Vivekanandha College of Arts and Sciences for Women.(Autonomous), Tiruchengode, India. Namakkal 637205.

Abstract—The Cloud Service Provider (CSP) provides application level services and storage level services to the users. The learners are interacted with the system using the user interface tool. All the trainer and learner information are managed under the cloud server. The trainer and learner communication tasks are carried out through the user interface application. The learning materials and evaluation documents are maintained under the storage in the cloud server. Authentication, confidentiality and data integrity operations are carried out under the security solutions. The storage data integrity is verified with public and private data verifiers. The cloud based e-learning system is constructed with privacy preserved public data verification mechanism. Third Party Auditors (TPA) are employed to verify the cloud storage data values. The multi user verification is achieved with batch and data dynamism auditing mechanism. The storage data privacy is ensured with Homomorphic linear authenticator and random masking techniques. The system is also protected with malicious TPA and server activities. All the data verification operations are carried out through the cloud resources. The storage services are improved with version management and traceability features.

Keywords— e-Learning Platforms, Trainers and Learners, Public Data Verification, Third Party Auditors and Batch Auditing Schemes

I. INTRODUCTION

The cloud computing model definition is introduced by National Institute of Standard Technology as “a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources can be rapidly provisioned and released with minimal management effort or services provider interaction.” Multi-tenancy and elasticity are two key characteristics of the cloud model. Multi-tenancy enables sharing the same services instance among different tenants. Elasticity enables scaling up and down resources allocated to a service based on the current service demand. Both characteristics focus on improving resources utilization, cost and service availability.

Cloud computing refers to both the application delivered as services over the Internet and the hardware and systems software in the data centers that provide those services. The services themselves have long been referred to as Software as a Services (SaaS). The data center hardware and software is what we will call a cloud. The US Department of Commerce’s National Institute of Standard and Technology defined cloud computing as :“a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources e.g. network, servers, storage, applications and services that can be rapidly provisioned and released with minimal management effort or service provider interaction.
Many trends are opening up the era of Cloud Computing, which is an Internet–based development and use of technology. Consider the large size of the outsourced electric data and the Client’s constrained resource capability, the core of the problem can be generalized as how can the client find an efficient way to perform periodical integrity verification without the local copy of data files. Cloud computing is a flexible way to allocate information technology(IT) resources i.e. storage, software, infrastructures and bandwidth etc. out of a pool, enabling to consume processing power according to user’s need. When there is a need to address peaks while saving costs when the users do not need the extra power any more. The global usage of a cloud leads to the optimization of resources so that in the end it makes them cheaper for everybody involved. The main objective of this paper is to provide security in terms of integrity and availability of client’s data which is stored on cloud. This paper shall also be taken care of by allowing trusted party to verify the correctness of the cloud data on demand without retrieving a copy of the whole data or introducing additional on-time burden to cloud users. Cloud Computing allows computer users to have a convenient access to fully featured applications, to software development and deployment environment and to computing infrastructure assets such as network-accessible data storage and processing.

The model is also well-suited for small and medium business it helps adopting IT without upfront investments in infrastructure, software licenses and other relevant requirement.

II. RELATED WORK

The paper is based on our previous work that describe the, E-learning course for scheduling of computer grid and design of algorithms and models to support the scheduling tasks in grid systems [8]. This paper is also partially related to the proposal of models for parallel tasks in workflow, designed in our previous work [10], [11].

The authors Frinkle and Morris [12] are oriented on HPC cluster building. The authors described a successful approach to designing and implementing a High Performance Computing (HPC) class in Southeastern Oklahoma State University. Class is focused on creating competency in building, configuring, programming, troubleshooting, and benchmarking HPC clusters. Students built three twelve-unit independently-operating clusters. Working groups were formed for each cluster and they installed the operating system, created users, connected to the campus network and wrote a variety of scripts and parallel programs while documenting the process [1]. Students gained substantial knowledge in fundamental aspects of HPC through the hands-on approach of creating their own clusters. Ping presents high performance computing environment to support research and teaching at a minority serving institution. The authors described teaching high-performance computing on a high-performance cluster and demonstrated the performance aspects 64-node Mozart cluster located at the University of Stuttgart. The authors introduced a program that combines education and research program. The program involves fundamental courses including high performance computing and parallelization. Good practices in teaching parallel computing and programming to science faculty are explained. The authors introduced the program of investigating principles of parallelism using people. Complex work describes the Bootable Cluster CD system running in RAM and having preconfigured environment and applications. The applications are on the base of MPI, PVM, openMosix, CUDA, C and other tools and languages. Authors discuss initiatives proposed by Louisiana State University to increase knowledge and understanding of high performance computing among high school students. The course contains lectures, assembly of equipment by students, interactive programming using PyMPI as well as tutorials. Czarnul [9] presents creation of the BeesyCluster environment. It is platform to make teaching and
access to HPC resources easy and allows MPI, OpenMP, CUDA-level application development as well as higher level service integration using a workflow management environment.

Brunner and Kim [2] describe an introductory data science course, entitled Introduction to Data Science, offered at the University of Illinois at Urbana-Champaign. The course introduced general programming concepts by using the Python programming language with an emphasis on data preparation, processing, and presentation.

### III. CLOUD BASED E-LEARNING APPLICATION AND PUBLIC DATA VERIFICATION SCHEMES

Cloud is on demand services using internet as medium and shares infrastructure with multiple end users in same resources. Cloud Application Service not required installation, fast deployment, greater scalability and cost optimization. It uses Cloud front for his distribution among multiple users. Course taught using social software like Application's tools the learner may have very few control over, but the academician or Administrator is the owner and has most of the control. Various number of parameter things in cloud based data, both applications’ functioning and resources are distantly available, accessible everywhere, salable assets.

Cloud storage services enable users to outsource their data to cloud servers and access the outsourced data remotely from a variety of places and devices. Such services provide users with an efficient and flexible way to manage their data without deploying and maintaining the local storage device and service [4], [5]. Some recent reports indicate that more than 79% of organizations attempt to utilize data outsourcing and such increasing demand of the cloud storage service leads to the growing number of cloud storage providers [10].

While people enjoy the desirable benefits from the cloud storage service, critical security concerns in data outsourcing have been raised seriously. One of the most important security concerns is data integrity. Since users do not physically own their data once outsourcing the data to cloud servers, they are always worried about the data integrity, i.e. whether their data remains intact on the cloud servers. The integrity check on users’ data can be performed by a cloud server, however the cloud server may always generate a good integrity report for good reputation even if some data are damaged or missing [8]. Users should be able to prevent the server from cheating. In addition, an external adversary may distort users’ data on the cloud servers for financial or political reasons [3]. As such, an efficient and secure verification method is often required by the users to ensure the integrity of their data.

Some data verification schemes rely on users themselves to execute the verification [6], [7]. However, this means that a user has to expend additional communication and computation costs to verify data integrity. As a consequence of this, the user is required to bear heavy communication and verification burden to retrieve and use the data. To reduce the verification burden on users, a public verification paradigm has been proposed. The idea is to employ an external and independent auditor to periodically verify the data integrity on behalf of users. As a result, users will be free from the verification burden while the auditor needs to be equipped with strong computation capability for verification. In existing public verification schemes, the computation overhead of verification by the auditor linearly increases with the size of the verified data set. If the verification is required to be executed frequently for multiple users’ data sets, the auditor will need a huge computation capability to accomplish the verifications and the verification delay will be huge. The deployment of such auditor is indeed a difficult problem. Therefore, to reduce the computation overhead and delay on the auditor side has a significant value to make the verification scheme efficient and practical.

Some public verification schemes achieve batch verification, where multiple delegated verification tasks from different users can be performed simultaneously by the auditor. However, the batch verification overhead in the current schemes still linearly increases with the number of users.
Consequently, if the auditor is equipped with a constrained device and the verifications are required to be executed frequently, the verification may incur a huge delay and become a bottleneck in applications. For example, for the public verification scheme, even though the auditor is equipped with an Intel Core 2 processor running at 1.86 GHz, 2,048 MB of RAM, let the size of the verified data set be 300, when the number of verification tasks (i.e. the number of users) increases to 100, the verification delay is about 30 seconds. And if the number of verification tasks increases to 1000, the verification delay is about 300 seconds.

IV. SECURITY ISSUES ON CLOUD BASED E-LEARNING PLATFORMS

The Cloud Service Provider (CSP) provides application level services and storage level services to the users. The learners are interacted with the system using the user interface tool. All the trainer and learner information are managed under the cloud server. The trainer and learner communication tasks are carried out through the user interface application. The learning materials and evaluation documents are maintained under the storage in the cloud server. Authentication, confidentiality and data integrity operations are carried out under the security solutions. The storage data integrity is verified with public and private data verifiers. Multi user based batch auditing scheme is employed for the data verification process. The following security issues are identified from the current cloud based e-learning platforms.

• Malicious auditors control is not supported
• Data security is not provided in the auditing process
• Data dynamism sequence is not monitored
• Privacy preservation is not supported in batch auditing process

V. PRIVACY PRESERVED PUBLIC DATA VERIFICATION SERVICES

The cloud based e-learning system is constructed with privacy preserved public data verification mechanism. The multi user verification is achieved with batch auditing and data dynamism mechanism. The storage data privacy is ensured with Homomorphic linear authenticator and random masking techniques. Version management, traceability and Malicious TPA control features are integrated with the system. The e-learning system is build to provide educational services with security and privacy. Batch auditing mechanism is adapted for the data verification process. Dynamic data updates are managed with auditing process. The system is divided into five major modules. They are Data center, Third Party Auditor, Client, Data dynamism handler and Batch auditing. The cloud data center manages the resources for the trainer and learner. Data verification operations are initiated by the Third Party Auditor. The trainer and learner communications are carried out with the client application. Data update operations are managed under data dynamism module. Batch auditing is designed for multi user data verification process.

The data center application is designed to allocate storage space for the trainer and learner. Data center maintains e-learning materials for multiple trainers. Different sized storage area is allocated for the trainers. Data files are delivered to the learners in the e-learning platform. The Third Party Auditor (TPA) maintains the signature for shared data files. TPA performs the public data verification for data providers. Data integrity verification is performed using Secure Hashing Algorithm (SHA). Homomorphic linear authenticator and random masking techniques are used for privacy preservation process.

The client application is designed to manage the communication between the trainers and learners. The shared data upload operations are carried out by the trainers. The shared data file download process is initiated by the trainers. Data access information is updated to the data center. Shared data values are managed with blocks. Block update and delete operations are handled with signature update
process. Block signatures are also updated in data dynamism process. Version management and traceability tasks are handled in the data dynamism handlers.

Data integrity verification is carried out under auditing process. Batch auditing is applied to perform simultaneous data verification process for multiple users. Data dynamism is integrated with batch auditing process. Cloud resources are used for the auditing process with malicious TPA handling mechanism.

VI. EXPERIMENTAL ANALYSIS

The cloud data audit scheme is designed for the multi user environment. The cloud data center manages the shared data for the users. The data owner uploads the data into the cloud data center. The Third Party Auditor (TPA) verifies the data values. The data verification is performed with privacy preserved manner. Cloud data center, Third Party Auditor (TPA) and client applications are developed in the system. He cloud data center manages the encrypted data values. The Third Party Auditor (TPA) handles the data integrity verification operations. The client application is designed to manage the data upload and download operations.

The cloud data sharing system is tested with two techniques. They are Privacy Preserved Batch Audit (PPBA) scheme and Privacy Preserved Batch Audit with Data Dynamism (PPBA-DD) scheme. The privacy preserved batch auditing scheme is used for the data verification with multi user support. The Privacy Preserved Batch Audit with Data Dynamism (PPBA-DD) scheme performs the data verification with data dynamism and malicious node detection features. The Enhanced PPBA-DD scheme support data traceability and version management in the system.

The system is tested with two performance parameters. They are response time and throughput measures. The response time is estimated with data requested time and data delivered time values. Figure 6.1 shows the response time analysis between Privacy Preserved Batch Audit (PPBA) scheme and Privacy Preserved Batch Audit with Data Dynamism (PPBA-DD) scheme. The response time in Privacy Preserved Batch Audit with Data Dynamism (PPBA-DD) scheme 25% reduced than the Privacy Preserved Batch Audit (PPBA) scheme. The throughput indicates the data delivery speed for the download process. Figure 6.2 shows the throughput analysis between Privacy Preserved Batch Audit (PPBA) scheme and Privacy Preserved Batch Audit with Data Dynamism (PPBA-DD) scheme. The Privacy Preserved Batch Audit with Data Dynamism (PPBA-DD) scheme increases the throughput 20% than the Privacy Preserved Batch Audit (PPBA).

Figure no.6.1 Response Time Analysis between Privacy Preserved Batch Audit (PPBA) scheme and Privacy Preserved Batch Audit with Data Dynamism (PPBA-DD) scheme
Figure no.6.2 Throughput Analysis between Privacy Preserved Batch Audit (PPBA) scheme and Privacy Preserved Batch Audit with Data Dynamism (PPBA-DD) scheme

VII. CONCLUSION

The e-learning applications are build with the support of the cloud resources. The Cloud Service Provider (CSP) act as an interface between the Trainer and Learner. Multi user based batch auditing scheme is applied for the data verification process. Privacy ensured public data verification scheme is build with data dynamism, version management, traceability and Malicious TPA control. The cloud based e-learning application is constructed with security and privacy. Public data verification is achieved with multi user based batch auditing scheme. Data dynamism is handled with version management and traceability features. Malicious activities are controlled with efficient resource utilization.

REFERENCES

[1] J. Skrinárova and E. Vesel, “Model of education and training strategy for the high performance computing”, IEEE, 2017.
[2] R. Brunner, E. Kim. Teaching Data Science. ICCS 2016. The International Conference on Computational Science, Procedia Computer Science, vol. 80, 2016, pp. 1947–1956
[3] Y. Zhang, C. Xu, J. Zhao, X. Zhang, and J. Wen, “Cryptanalysis of an integrity checking scheme for cloud data sharing,” Journal of Information Security and Applications, vol. 23, pp. 68–73, 2015.
[4] Y. Cui, Z. Lai, X. Wang, N. Dai and C. Miao, “QuickSync: Improving synchronization efficiency for mobile cloud storage services,” in Proceedings of MobiCom. ACM, 2015, pp. 592–603.
[5] H. Li, Y. Yang, T. H. Luan, X. Liang, L. Zhou and X. Shen, “Enabling fine-grained multi-keyword search supporting classified sub-dictionaries over encrypted cloud data,” IEEE Transactions on Dependable and Secure Computing, vol. 13, no. 3, pp. 312–325, 2016.
[6] D. Cash, A. K‘upe, ‘u, and D. Wichs, “Dynamic proofs of retrievability via oblivious ram,” in Proceedings of EUROCRYPT. Springer, 2013, pp. 279–295.
[7] E. Shi, E. Stefanov, and C. Papamanthou, “Practical dynamic proofs of retrievability,” in Proceedings of CCS. ACM, 2013, pp. 325–336.
[8] J. Skrinarova, L. Huraj, V. Siladi, “A neural tree model for classification of computing grid resources using PSO tasks scheduling,” in Neural networks world. 2013. ISSN 1210-0552
[9] P. Czarnul. Teaching High Performance Computing using BeesyCluster and Relevant Usage Statistics, ICCS 2014. 14th International Conference on Computational Science, Procedia Computer Science, Volume 29, 2014, pp. 1458–1467.
[10] J. Skrinarova., M. Povinsky, GPGPU based job scheduling simulator for hybrid high-performance computing systems. Edited by: Novitzka, V; Korecko, S; Szakal, A Conference: IEEE 13th International Scientific Conference on Informatics, 2015, Location: Poprad, SLOVAKIA Date: NOV 18-20, 2015.
[11] J. Skrinarova, “Implementation and evaluation of scheduling algorithm based on PSO HC for elastic cluster criteria,” in Central European Journal of computer science. 4 (3) 2014, pp. 191-201.
[12] K. Frinkle, M.Morris, Developing a Hands-On Course Around Building and Testing High Performance Computing Clusters, ICCS 2015 International Conference On Computational Science, volume 51, 2015, pp. Pages 1907–1916