Cluster Computing (2019) 22:S14999–S15006
https://doi.org/10.1007/s10586-018-2480-9

Efficient data integrity and data replication in cloud using stochastic diffusion method

M. Ramanan[1] · P. Vivekanandan[1]

Received: 31 January 2018 / Revised: 5 March 2018 / Accepted: 8 March 2018 / Published online: 15 March 2018

Abstract
Cloud computing will provide scalable computing as well as storage resources where more data intensive applications will be developed in a computing environment. Owing to the existence of such security threats in the cloud, several mechanisms are being proposed for allowing the users to audit the integrity of data along with the public key of the owner of the data even before making use of the cloud data. Replicating of data in cloud servers through multiple data centers offers better availability, scalability, and durability. The correctness of choice of the right type of public key of the previous mechanisms is based on the security of the public key infrastructure (PKI). Although traditional PKI has been widely used in the construction of public key cryptography, it still faces many security risks, especially in the aspect of managing certificates. There are different applications having different types of quality of service (QoS) needs. In order to support the QoS requirement continuously, the application of such data corruption for this work will be an efficient integrity of data replication that makes use of a stochastic diffusion search (SDS) algorithm that has been proposed. This SDS is that technique of a multi-agent global optimisation which has been based on the behaviour of ants that has been rooted in the partial evaluation of that of an objective function along with direct communication among agents. The proposed SDS algorithm will minimize the replication cost of data. The results of these experiments have shown that the mechanism will be able to demonstrate the effectiveness of this proposed algorithm which is in the replication of data as well as its recovery. The proposed method when appropriately compared with the cost effective replication of dynamic data given by Li et al. proves that the average recovery time is less by 18.18% for the 250 number of requested nodes, by 14.28% for the 500 number of requested nodes, by 11.11% for the 750 number of requested nodes and by 8.69% for the 1000 number of requested nodes.

Keywords Cloud computing · Cloud services · Data storage · Data integrity · Data replication and stochastic diffusion search (SDS)

1 Introduction

Cloud computing is that emerging technology that has got a lot of attention recently for providing services on the internet. The users will be able to use the online services for various software as opposed to that of purchasing or even installing them on their computers. The National Institute of Standard and Technology (NIST) have defined cloud computing as that paradigm that enables the useful and on demand access to the network to that of a shared pool of the configurable resources. It further offers many other services that are presented in three different models which are: the software as service (SaaS), the platform as service (PaaS), and the infrastructure as service (IaaS). The software as service (SaaS) further provides services that exist in cloud or in the applications to end users, the platform as service (PaaS) will provide access to the platforms and the infrastructure as service (IaaS) provides the processing storage and also the other computing resources [1].

✉ M. Ramanan
proframanan.cse@gmail.com; pmramanan@gmail.com

P. Vivekanandan
anandpvivek@yahoo.co.in

[1] Department of CSE, Park College of Engineering and Technology, Kaniyur, Coimbatore, India

Security is perhaps the main barrier to continuing growth of cloud computing and for the purpose of certain security risks and some issues, the enterprises and the individuals are not willing to deploy data as well as applications in the cloud environment. When the data, the web applications, and the services get hosted in a cloud environment by the providers of service, the control of all of these is not managed by them. The cloud services are the shared infrastructure for increasing the vulnerabilities relating to access of unauthorised data concerning privacy of data, identity management, compliance, authentication, confidentiality, availability, encryption and internet protocol (IP) vulnerabilities (which in most cases will be untrusted that permits a man in the middle of the attack), the network security as well as physical security [2].

Generally, the cloud computing will provide both software and hardware services that use large scale data centres. The result of this was that the cloud computing moved further away from the data storage and computation. From that of the end user onto the servers that are located within the data centres. Thereby relieving the users from the hardship of both application provisioning and management. As a result, software can then be thought of as purely a service that is delivered and consumed over the Internet. Offering users the flexibility to choose applications on-demand and allowing providers to scale out their capacity accordingly. It has been quite challenging to provide a high level of availability and also an efficient access to the data centres owing to the large scale as well as its dynamic nature [3].

The increase in the volume of both personal and vital data will bring further focus on the storing of this data in a secure manner. By implementing the services of cloud computing, the local storage and its reliance in addition to bringing down the maintenance and operations costs are observed. The users will have a major security as well as privacy concerns on their outsourced data owing to the possible unauthorised access in the service providers. Darwazeh et al. [8] made a proposal for a secure cloud computing model that is based on the classification of data. This proposed cloud model will bring down the overhead and the time for processing for securing the data by using various security mechanisms having different sizes of the key for providing proper confidentiality levels. This model had been tested having a different algorithm for encryption and the results of simulations have shown the reliability as well as the efficiency of this proposed framework.

Fabian et al. [9] have presented another novel architecture and the implementation of the data sharing that is inter-organizational that gives a high level of privacy and security for the patient data in the cloud computing environments. The architecture features an encryption that is attribute-based for reduction of the adversarial capacity of the cloud computers that are curious. The implementation, as well as evaluation by means of many experiments, will demonstrate the feasibility as well as a good performance of this approach.

Jiang et al. [10] brought about a collision attack for the current scheme and also provided a public integrity auditing scheme having a secure group user revocation that has been based on the vector commitment as well as the verifier-local revocation group signature. They further designed one more concrete scheme that was based on the definition of the scheme. This supports the public checking as well as effective revocation of the user with some nice properties like efficiency, traceability, and confidentiality. Lastly, the experimental analysis and security when compared with the other relevant schemes, this particular work scheme has also been secure and efficient.

Data owners store their data remotely in the cloud and have access to on-demand high quality applications and services. However, data owners and cloud servers are not in the same trusted domain, due to which the data's integrity may be at risk. Generally, owners or a trusted third party audit data storage. Cloud service providers (CSP), to cut costs tend to discard some lesser accessed data or delegate it to second-level storage devices. Malicious CSPs may delete data or sell the data to competitors. Also, hackers may intercept and capture the communications and access user's sensitive information. Thus, processes to verify the data's confidentiality and integrity is required. Users require assurances that their data is secure. Thus, an efficient and secure scheme for cloud data storage is essential for ensuring the data integrity and confidentiality.

Data replication is a commonly used technique for increasing the data availability in cloud computing. Cloud replicates the data and store them strategically on multiple servers located at different geographic locations. A replication is that process of getting different replicas of a similar service on different nodes.

Replication is a used technique in different clouds, such as google file system (GFS) and hadoop distributed file system (HDFS). In case of the cloud, the data replication will be achieved using a data resource pool along with the actual number of data replicas that is set based on both history and experience. Also, it will not be needed to create another replica for all of the data files. So it becomes necessary to replicate adaptively the popular files and determine the actual number of replicas along with the data nodes and where the replicas have to be placed based on the current environmental conditions of the cloud [4].

Ali et al. [11] made a proposal of a division and replication of data in that of the cloud for the purpose of optimal performance and security (DROPS) that will collectively approach this security and also the issues in performance.

In case of the DROPS methodology, it can divide the file into different fragments, and further replicate this fragmented data over that of the cloud nodes. Each node stores only a small or a single fragment of a certain data file ensuring that even in the case of any successful attack, no type of meaningful information will be exposed to this attacker. Furthermore, these nodes that store fragments, are being separated with that of a particular distance by means of the graph T-colouring for prohibiting attackers to guess the locations of such fragments. Also, this DROPS methodology will not depend on the techniques of traditional cryptography for data and its security thus relieving this system from the methodologies that are expensive in terms of computation.

Zhang et al. [12] made a proposal of provable multiple replication data possession protocol having full dynamics called the MR-DPDP. In case of the MR-DPDP, it will make use of an authenticated data structure that is called the Merkle hash tree that has a rank for supporting both the dynamic data updates and the efficient integrity and its verification. Additionally, construction with that of the Ron Rivest, the Adi Shamir and the Leonard Adleman (RSA) signature will support the variable-sized file blocks and using security proof along with performance evaluation this will demonstrate the MR-DPDP for incurring a lower overhead of communication while updating the data blocks and also verifying the proof of integrity for many more replicas.

A cloud storage service is a very common as well as popular service (for example the Google Drive, the Dropbox, the Amazon S3 and the microsoft one drive) for usage by general users. However, the users have also faced a bottleneck on its local side storage space as they need a large storage space for storing a large amount of data on this type of a situation. The cloud storage service will have a very high capacity and also a high computation solving users' difficult problems. Moreover, the user tends to build a larger storage device that may be more expensive than that of the rented cloud storage service. As the cloud storage service provides an access to cloud services from the web service or the applications using the application programming interface (API) by means of mobile devices (like the laptop, the table computer, and the smart phones), it is very convenient to be used by the users, and also achieves ubiquitous service [5].

Although any cloud storage service has several advantages, it can bring many challenging issues that include efficacy or security. One such challenge is the verification of integrity as users will not know how this storage service can handle data. Such cloud storage services will be provided by the commercial enterprises to ensure they are not fully trusted. So the provider can also hide the data loss or the data errors in service owing to their benefits. It is a traditional approach by downloading the whole data from cloud and verify its integrity by means of checking the digital signatures or the hash values of the entire data. This type of a simple approach will be able to check the integrity of the data users. Owing to a large size of such data that is outsourced and the limited capability of the users, there needs to be found an efficient way for achieving the verifications of integrity without any local data file copy. To solve the issue of data integrity verification several investigations that are present in different methods, as well as security models, are employed [6].

The existing public schemes of verification will assume the auditor being honest and will not be corrupted and this being a strong assumption as auditors may be corrupt in practice. There may be a malicious auditor that can claim outsourced data which is not well retained and additionally the vulnerability of the current schemes may be further exacerbated by the malicious auditors colluding with these cloud servers and also generate a message that is biased and challenging for checking the data blocks which are not corrupt and so will deceive the users. The construction of an efficient verification of the data integrity of the cloud storage against such malicious auditors will be of great importance [7].

Encryption of the data before storing in cloud addresses the confidentiality issue. Though, verifying integrity of data is a hard task without a local copy of data. Thus, cryptographic primitives cannot be used for protecting outsourced data. Downloading of data for cross checking its integrity is unfeasible due to high I/O cost, high communication overhead and limited computing capability. Therefore, efficient and effective mechanisms are required for protecting the confidentiality and integrity of user's data with minimal computation, communication and storage overhead.

This advanced network explorations with the growing demand for sharing and transferring of mobile data that has driven many novel applications in that of the cyber-physical systems (CPSs), like the intelligent transportation systems (ITSs). The current implementations of the ITS that are restricted by means of conflicts among communication efficiency and security. Based on this issue, Gai et al. [13] made a proposal of a security-aware efficient data sharing along with transferring (SA-EAST) model, that has been designed for the purpose of securing the implementations. While applying this approach they aim at obtaining a secure and a real time transferring and multimedia data sharing. This evaluation has proved that such models will provide one effective performance in terms of securing the communications of the ITS.

Sookhak et al. [14] made a proposal for another efficient remote data auditing (RDA) technique that is based on the algebraic signature properties for that of a system of cloud

storage which will incur a minimum computational as well as communication costs. This also made a presentation of a new data structure called the divide and conquer table (DCT) which supported the dynamic operations of data like the append, the insert, the modify, and delete. This proposed data structure may also be applied for that of a data storage of a large-scale. When compared with the RDA techniques this method will be secure and be highly efficient in the reduction of the costs of communication.

Dashti and Rahmani [15] made a proposal for an architecture that satisfied both the consumers as well as the providers in terms of the needs of this technology. They further designed another new service within the PaaS layer for the scheduling of the tasks of consumers. Incompatibility between the specification of such physical machines and the user requests in the cloud will result in problems like the trade-off and large consumption of power to decrease profits. For this, the QoS of the users and the reduction of energy has been proposed by using the particle swarm optimization (PSO) in order to reallocate migrated virtual machines in that of the overloaded host. They further dynamically consolidated the under-loaded host that provides power saving.

Li et al. [16] also developed another dynamic energy-efficient virtual machine (VM) migration along with a consolidation algorithm based on the models of multi-resource energy-efficient. Here the authors also designed the method of double threshold having a multi resource use for triggering the VMs and their migration. This modified PSO method has been introduced into this consolidation of the VMs for avoiding them from falling into the local optima. When compared with the other heuristic algorithm this is modified as best fit decrease and reduced the actual number of such physical nodes and shows better degrees of efficiency in case of a data centre for cloud computing.

Lin and Chong [17] further presented another genetic algorithm (GA) that was based resource constraint project scheduling by incorporating various new ideas (the enhancements as well as the local search) for the purpose of solving computing and resources allocation problems in that of a cloud manufacturing system. The generated offspring will not be feasible owing to the precedence of the task and the constraints of resource availability. There are conflict resolutions as well as enhancements that have been performed in this and the neighbourhood of solutions can be exploited and owing to the complex traits the allocation of computing resources in a system of cloud manufacturing is found to be NP-hard. The results of computation have shown that this proposed GA will be able to provide a better schedule which will be able to allocate the computing resources optimally.

Some existing public verification schemes cannot resist a common attack where an external, active and online adversary can modify the cloud stored data and tamper with the interaction messages between the cloud server and the auditor. The proposed scheme is secure against such external adversary. One common approach for resisting these adversaries will be to have this cloud server interact with auditors for securing a channel. The construction of a secure channel for each of the tasks is a cumbersome task.

In this paper, mechanism for maintaining the integrity of data and the replication in that of the cloud computing is proposed. The proposed method is based on the SDS algorithm. The rest of the text has been organized as follows: in Sect. 2 several methods used in the work are discussed. In Sect. 3 the experimental results are observed and the work is concluded in Sect. 4.

2 Methodology

This QoS aware data replication (QADR) problem is concerned about how the QoS needs can be efficiently considered. The goal of this QADR will be to bring down the cost of data replication. By minimising this, the probability of data corruption is reduced significantly. Owing to the limited space of replication in a storage node, the replicas of data in certain applications can be stored in the nodes of lower performance. This can lead to some replicas which will not meet the requirements of the QoS and their applications. These replicas are the QoS-violated data replicas and the number of such violated replicas will be expected to be very small. For solving this QADR problem, the work has proposed an SDS algorithm.

2.1 Data integrity

As according to Fig. 1, this system model includes four different entities: the cloud, the data owner, data users and finally the Key Generation Centre (KGC). Cloud further provides data services to both the owner and the user. The owner will outsource data to the cloud and save them on local devices. Generally, for efficient modification, this data will be divided into different blocks. The data user will use the cloud data that was outsourced by the data owner in the cloud. The data user also performs a search on the cloud data for certain purposes. This KGC will be a trusted party in this framework. The KGC is that trusted party that is needed within the framework of the certificate-less schemes which can generate a partial and private key for the entity (the data owner) and based on its identity (name or email address) and the rest of the private key will be generated by this entity itself [18].

The data that is stored in the cloud can also be polluted from two different possible causes. Firstly, an external adversary can pollute data, and also prevent the owner as

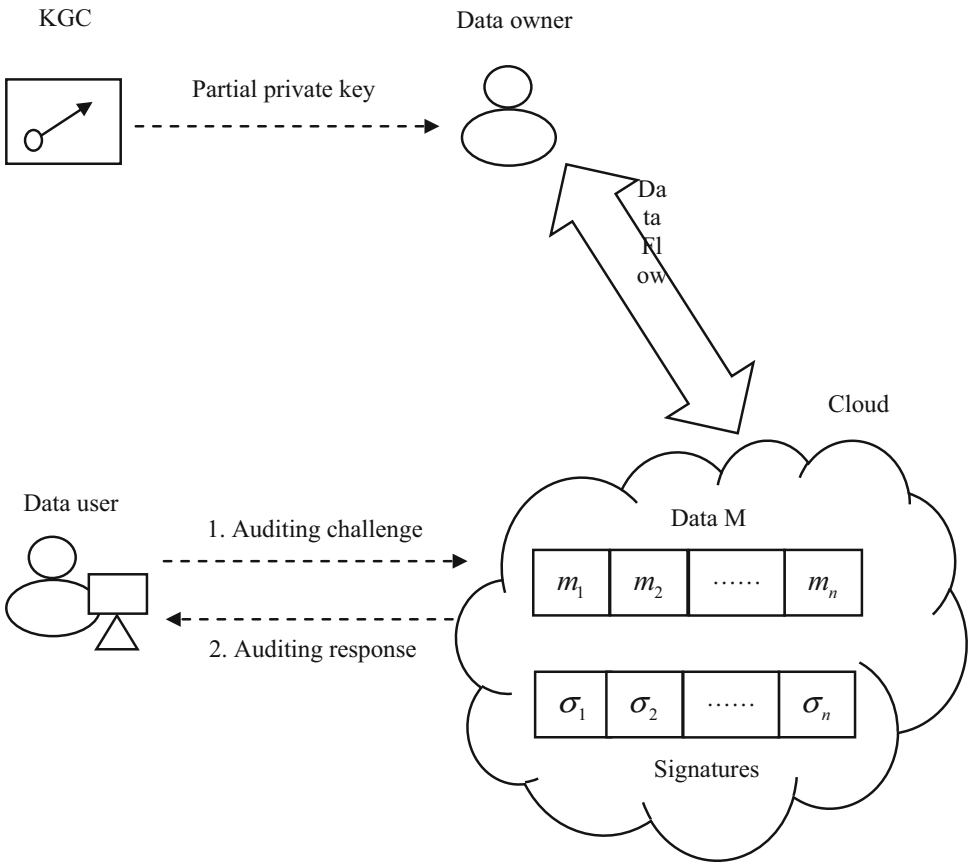

Fig. 1 The cloud, the data owner, data users and the KGC

well as the users from making use of this data accurately. Secondly, the providers of cloud service may also corrupt the data accidentally due to human errors. Therefore, the owners and the users may not be able to fully trust the cloud.

For the purpose of protecting the integrity of data which has a signature attached to the private key of the owner. The data user has to check on the integrity of the cloud data like the search, the computation and the data mining after which the cloud will generate a proof of possession. Lastly, the data user will verify the integrity that is based on the response of auditing. It has to be noted that the owner will also have to be a verifier for checking the data and its integrity that is not possessed physically by means of following this protocol.

This design of such a public auditing mechanism will get three objectives which are: (1) Correctness: a public verifier (i.e., a data user) is able to verify the integrity of data in the cloud correctly. (2) Public auditing: a public verifier is able to audit the correctness of data without retrieving the entire data from the cloud and (3) Certificate less: which denotes the correctness of such public auditing which will not need a public verifier for managing certificates.

2.2 Stochastic diffusion search (SDS)

The SDS will be based on the computation that is distributed where the operating of simple units or the agents that are probabilistic is inherent. There is a positive feedback that promotes some better solutions by means of allocating more agents to them for exploration. The limited resources will induce a strong competition from one of the larger population of agents that correspond to the solution that is best-fit which emerges rapidly [19].

In many of such search problems, solutions are thought of as being composed of several results that unlike many methods of swarm intelligence, the SDS will use this decomposition for increasing the efficiency of search. Here, each agent will pose a hypothesis on the possible solution which will evaluate this partially. The successful agents will continue to test this hypothesis at the same time recruiting the unsuccessful agents using a direct communication. This can bring about a mechanism of a positive feedback to ensure a rapid convergence of the agents to promising solutions in this space. There is a global solution thus constructed from the interaction of several simple agents that form a large cluster. This cluster will be a large cluster which is dynamic in nature and will be stable and

also analogous to that of "a forest whose contour does not change but whose individual trees do change".

Unlike many such nature inspired search algorithms, the SDS [20] also has a strong framework of mathematics that describes the behaviour of this algorithm by means of investigating the resource allocation, the convergence to global optimum, the robustness and finally the minimal convergence criteria along with the linear time complexity. This pseudo code of the SDS algorithm has been given below:

Initialising agents()
While (*stopping condition is not met*)
Testing hypotheses ()
Diffusion hypotheses ()
End

During the phase of initialisation, all the agents will select randomly a hypothesis from a search space. These agents are all set to be inactive. Two types of knowledge that can influence this initialisation phase: (i) The actual ratio of the size of this model to the size of search space is greater than one; this will guarantee that at least one single agent is duly initialised with one of the best hypothesis. (ii) The earlier location of this model will be known; this will be useful at the time of performing successive searches on such similar search spaces; for instance, the consecutive frames of a video.

During this test phase, these agents determine whether it will have to set itself to be more active or inactive. This has been achieved by means of applying a single test function to that of its current hypothesis. This type of a test function is also a partial evaluation of the position of the hypothesis. The test function also differs depending on the domain of application. The agents have been set to active in case this partial evaluation of hypothesis will return success; else, they continue to be inactive [21].

During the phase of diffusion, the agents exchange information on the hypothesis. The idea for this is the active agents that disseminate their hypothesis to the inactive agents. Three differing strategies in dissemination are identified, called recruitment strategies: the passive recruitment, the active recruitment and a combination of both. This type of an information exchange will lead to proper hypotheses in recruiting inactive agents and a large number of such agents will congregate around this hypothesis that is available. A standard SDS strategy of recruitment will be deployed and passive.

A relate phase will be an optional phase that is introduced in case there are multiple models that are extant within the search space. The technique permits a dynamic re-allocation of the agents. This relate phase will also help in a dynamic search space for re-aligning themselves using the right hypotheses. A relate phase will have two modes: the context free and the context sensitive.

The weak halting criteria will state that the SDS needs to stop when a percentage of all the agents are active, regardless of the hypothesis. There is a stabilisation that is seen as an active agent population that is steady with some margin of tolerance. Once this is met, the search will stop. The strong criteria of halting will define the halt state as related to the percentage of the active agents in a large cluster by applying the threshold or the tolerance rule as the weak halting state but instead looking at the percentage of such agents that are active inside this large cluster.

$$\text{Minimize}\left(\sum_{\forall r_i \in S_r}\sum_{\forall q_i \in S_n^{\overline{R(r_i)}}} x(r_i, q_j) \times T_{\text{storage}}(r_i, q_j)\right) + \left(\sum_{\forall r_i \in S_r}\sum_{\forall q_i \in S_n^{\overline{R(r_i)}}} y((r_i, q_j) \times k\right) \quad (1)$$

In the Eq. (1), the main objective will be to minimize the first minimum term which is the total replication cost of all the data replicas, the second minimum term will be the actual number of such QoS-violated data replicas. The next minimum term will be prior to that of the first minimum term. A coefficient k will be used for ensuring that the actual number of the QoS violated data replicas will be among the first minimized [22]. The main reason will be explained subsequently. In the Eq. (1), in case the requested node r_i will put one data block replica within the node q_j, this particular event will be recorded by means of setting 1 in $x(r_i, q_j)$. However, in case this replica is one of a QoS-violated data block type of a replica, $y(r_i, q_j)$ it will also be set as 1. By means of adding all these values of y, the number of QoS violated data replicas will be obtained. This number will be expected to be small by means of associating with one constant coefficient $k = \max\forall r_i \in S_r \wedge \forall q_j \in S\{T_{storage}(r_i, q_j)\} + 1$. With the actual setting of k, each $y(r_i, q_j)$ has a larger coefficient than $x(r_i, q_j)$. This is called the values of the $x(r_i, q_j)$ and $y(r_i, q_j)$ which are either 0 or 1.

3 Results and discussion

In this section, the random, cost effective dynamic data replication (Li et al.) and proposed methods are used. The worst case recovery time, average recovery time, recovery time after 20% data corruption and recovery time after 40% data corruption are shown in Figs. 2, 3, 4 and 5.

From the Fig. 2, it can be observed that the proposed method has lower worst case recovery time by 25.05 and

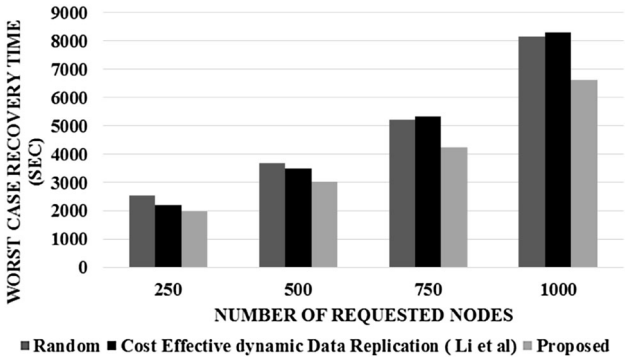

Fig. 2 Worst case recovery time (second)

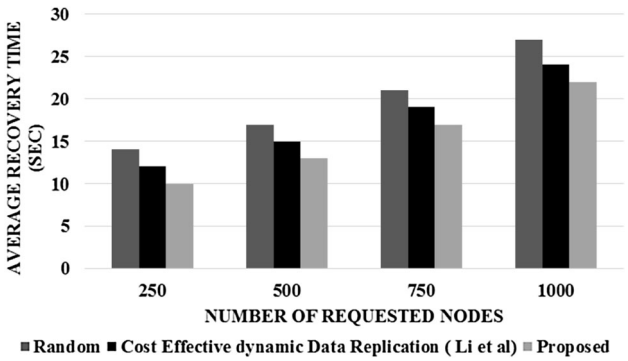

Fig. 3 Average recovery time (second)

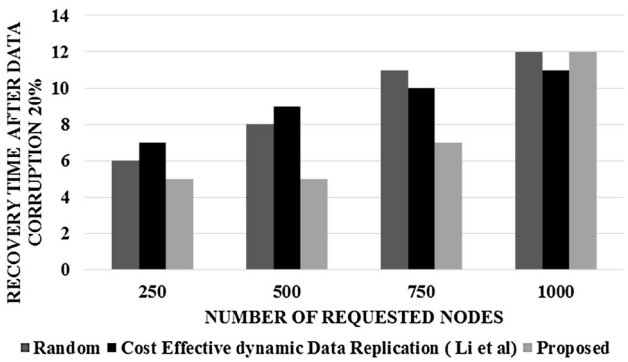

Fig. 4 Recovery time after data corruption 20%

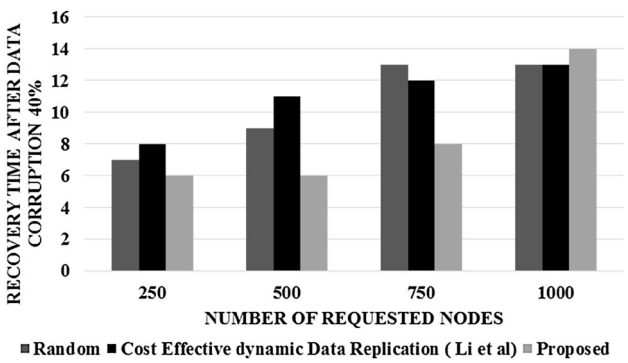

Fig. 5 Recovery time after data corruption 40%

11.36% for 250 number of requested nodes, by 19.24 and 13.9% for 500 number of requested nodes, by 20.16 and 22.63% for 750 number of requested nodes and by 20.75 and 22.86% for 1000 number of requested nodes when compared with random and cost effective.

From the Fig. 3, it can be observed that the proposed method has lower average recovery time by 33.33 and 18.18% for 250 number of requested nodes, by 26.66 and 14.28% for 500 number of requested nodes, by 21.05 and 11.11% for 750 number of requested nodes and by 20.4 and 8.69% for 1000 number of requested nodes when compared with random and cost effective dynamic data replication (Li et al.).

From the Fig. 4, it can be observed that the proposed method has lower recovery time after data corruption 20% by 33.33 and 18.18% for 250 number of requested nodes, by 46.15 and 57.14% for 500 number of requested nodes, by 44.44 and 35.29% for 750 number of requested nodes, but the proposed method has higher recovery time after data corruption 20% by same value and 8.69% for 1000 number of requested nodes when compared with random and cost effective dynamic data replication (Li et al.).

From the Fig. 5, it can be observed that the proposed method has lower recovery time after data corruption 40% by 15.38 and 28.57% for 250 number of requested nodes, by 40 and 58.82% for 500 number of requested nodes, by 47.61 and 40% for 750 number of requested nodes, but the proposed method has higher recovery time after data corruption 40% by 7.4 and 7.4% for 1000 number of requested nodes when compared with random and cost effective dynamic data replication (Li et al.).

4 Conclusion

In case of the cloud storage service, the integrity of data of the verification is a very critical issue. Here in this work, the organizing of public auditing requirements that contain function, security as well as the performance from many relevant kind of literature proposes the SDS method which is an optimization algorithm based on the interaction of agents. There is a high level description of the SDS shown as a social metaphor that demonstrates the procedures of SDS allocation of resources. This SDS algorithm will achieve an optimal solution to the problem of QADR. The cost of data replication and the QoS-aware data replicas are minimized. The results show that this proposed method has a lower average recovery time by about 33.33 and 18.18% for the 250 number of requested nodes, by about 26.66 and 14.28% for the 500 number of requested nodes, by about 21.05 and 11.11% for the 750 number of requested nodes and by about 20.4 and 8.69% for the 1000 number of

requested nodes when duly compared with the random and the cost effective replication of dynamic data (Li et al.).

References

1. Jouini, M., Rabai, L.B.A.: A security framework for secure cloud computing environments. Int. J. Cloud Appl. Comput. **6**(3), 32–44 (2016)
2. Zunnurhain, K., Vrbsky, S. V.: Security in cloud computing. In: Proceedings of the 2011 International Conference on Security and Management (2011)
3. Hussein, M.K., Mousa, M.H.: A light-weight data replication for cloud data centers environment. Int. J. Eng. Innov. Technol. **1**(6), 169–175 (2012)
4. Boru, D., Kliazovich, D., Granelli, F., Bouvry, P., Zomaya, A.Y.: Energy-efficient data replication in cloud computing datacenters. Clust. Comput. **18**(1), 385–402 (2015)
5. Liu, C.W., Hsien, W.F., Yang, C.C., Hwang, M.S.: A survey of public auditing for shared data storage with user revocation in cloud computing. IJ Netw. Secur. **18**(4), 650–666 (2016)
6. Zhang, J., Dong, Q.: Efficient ID-based public auditing for the outsourced data in cloud storage. Inf. Sci. **343**, 1–14 (2016)
7. Zhang, Y., Xu, C., Li, H., Liang, X.: Cryptographic public verification of data integrity for cloud storage systems. IEEE Cloud Comput. **3**(5), 44–52 (2016)
8. Darwazeh, N.S., Al-Qassas, R.S., AlDosari, F.: A secure cloud computing model based on data classification. Proc. Comput. Sci. **52**, 1153–1158 (2015)
9. Fabian, B., Ermakova, T., Junghanns, P.: Collaborative and secure sharing of healthcare data in multi-clouds. Inf. Syst. **48**, 132–150 (2015)
10. Jiang, T., Chen, X., Ma, J.: Public integrity auditing for shared dynamic cloud data with group user revocation. IEEE Trans. Comput. **65**(8), 2363–2373 (2016)
11. Ali, M., Bilal, K., Khan, S., Veeravalli, B., Li, K., Zomaya, A.: DROPS: division and replication of data in the cloud for optimal performance and security. In: IEEE Transactions on Cloud computing (2015)
12. Zhang, Y., Ni, J., Tao, X., Wang, Y., Yu, Y.: Provable multiple replication data possession with full dynamics for secure cloud storage. Concurr. Comput.: Pract. Exp. **28**(4), 1161–1173 (2016)
13. Gai, K., Qiu, L., Chen, M., Zhao, H., Qiu, M.: SA-EAST: security-aware efficient data transmission for ITS in mobile heterogeneous cloud computing. ACM Trans. Embed. Comput. Syst. **16**(2), 60 (2017)
14. Sookhak, M., Gani, A., Khan, M.K., Buyya, R.: Dynamic remote data auditing for securing big data storage in cloud computing. Inf. Sci. **380**, 101–116 (2017)
15. Dashti, S.E., Rahmani, A.M.: Dynamic VMs placement for energy efficiency by PSO in cloud computing. J. Exp. Theor. Artif. Intell. **28**(1–2), 97–112 (2016)
16. Li, H., Zhu, G., Cui, C., Tang, H., Dou, Y., He, C.: Energy-efficient migration and consolidation algorithm of virtual machines in data centers for cloud computing. Computing **98**(3), 303–317 (2016)
17. Lin, Y.K., Chong, C.S.: Fast GA-based project scheduling for computing resources allocation in a cloud manufacturing system. J. Intell. Manuf. **28**(5), 1189–1201 (2017)
18. Wang, B., Li, B., Li, H., Li, F.: Certificateless public auditing for data integrity in the cloud. In: 2013 IEEE Conference on Communications and Network Security (CNS), pp. 136–144. IEEE (2013)
19. Williams, H., Bishop, M.: Stochastic diffusion search: a comparison of swarm intelligence parameter estimation algorithms with ransac. Algorithms **7**(2), 206–228 (2014)
20. El-henawy, I.M., Ismail, M.M.: A hybrid swarm intelligence technique for solving integer multi-objective problems. Int. J. Comput. Appl. (2014). https://doi.org/10.5120/15192-3571
21. Al-Rifaie, M. M., Bishop, M. J., Blackwell, T.: An investigation into the merger of stochastic diffusion search and particle swarm optimisation. In: Proceedings of the 13th annual conference on Genetic and evolutionary computation, pp. 37–44. ACM (2011)
22. Lin, J.W., Chen, C.H., Chang, J.M.: QoS-aware data replication for data-intensive applications in cloud computing systems. IEEE Trans. Cloud Comput. **1**(1), 101–115 (2013)

M. Ramanan received his B.Tech (Information Technology) and M.E. (Computer Science and Engineering) from Anna University, Chennai. He has 5 years of teaching experience. He is currently working as Assistant Professor in the Department of Computer Science and Engineering at Park College of Engineering and Technology, Coimbatore, India. He has published many research papers in National/International Conferences and Journals. His research interest is mainly focused on managing the integrity of data in Cloud Storage environment.

P. Vivekanandan is currently working as a Professor, Department of Computer Science and Engineering, Park College of Engineering and Technology, Coimbatore, Tamilnadu, India. He has more than 12 years of teaching experience. He obtained his B.E. (Computer Science and Engineering) from Bharathiar University, Coimbatore, India and his M.Tech (Distributed Computing Systems) from Pondicherry University, Pondicherry, India and his Ph.D. from Anna University Chennai. At present, He is guiding 10 research scholars of Anna University, India. His research interests include Knowledge Discovery and Data Mining, Soft Computing and Distributed Computing. He has published many research papers in National/International Conferences and Journals. He has attended several seminars and workshops in the past 10 years. He has also organized several symposiums and workshops. He has guided more than 20 UG and 15 PG projects. He is a life member of ISTE and also a member of Computer Society of India.