# A SECURE ALGORITHM FOR DATA DEDUPLICATION IN CLOUD STORAGE SYSTEM

ANKIT BARSKAR,
Department of Electrical & Electronics Engineering,
Truba Institute of Engineering & Information Technology, Bhopal, India

NEPAL BARSKAR,
Department of MCA, University Institute of Technology RGPV, India

## ABSTRACT

Data redundancy is a significant issue that wastes plenty of storage space in the cloud-fog storage integrated environments. Most of the current techniques, which mainly center around the static scenes, for example, the backup and archive systems, are not appropriate because of the dynamic nature of data in the cloud or integrated cloud environments. This problem can be effectively reduced and successfully managed by data deduplication techniques, eliminating duplicate data in cloud storage systems. Implementation of data deduplication (DD) over encrypted data is always a significant challenge in integrated cloud-fog storage and computing environment to optimize the storage efficiently in a highly secured manner. This paper develops a new method using AES, Blowfish and RSA algorithms over the cloud and fog environment to construct secure deduplication systems. The proposed method focuses on the two most important goals of such systems. On one side, the redundancy of data needs to be reduced to its minimum, and on the other hand, a robust encryption approach must be developed to ensure the security of the data. Our approach found the advantage of the RSA encryption algorithm over the existing Attribute-based encryption. Which makes the advantage of symmetric key encryption and performed the fast process over the existing work scenario to prove the efficiency of our algorithm? Our work is computed using the parameter computation time which is efficient than the existing attribute-based encryption algorithm.

## 1. INTRODUCTION

Introduction Cloud enrolling is a flow mechanical headway in the handling field in which generally revolved around delineating of organizations given customers a similar course central sustenance, power correspondence. Advancement organizations made encouraged (a framework proposed securing datacenter) afterward organizations customers reliably at whatever point they have. Facilitated administrations conveyed clients utilize, occupancy, versatility, request a practical way. Distributed computing is turned out to be prominent as a result of the above say administrations clients. Organizations customers are given expert center (Internet pro association) web figuring. Web development, inventive headway in and scattered figuring and getting to of the fast associate insignificant exertion the focal point of customer innovation. Advancement is structured thought of organization customers without acquiring organization set away adjacent. [26]

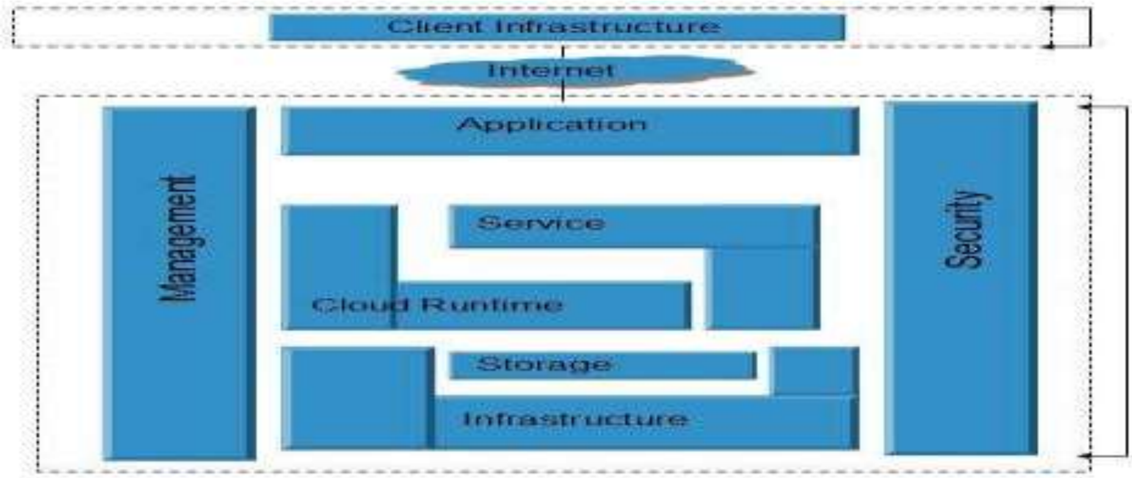

**Figure1.1: Cloud Architecture**

### 1.1 Cloud Service Models

The organizations gave by the disseminated processing are isolated into three all-around recognized classes. On a very basic level, organization one another and arranged 3-levels designing. Figure 1.2 shows the 3-level structure of dispersed processing[24].

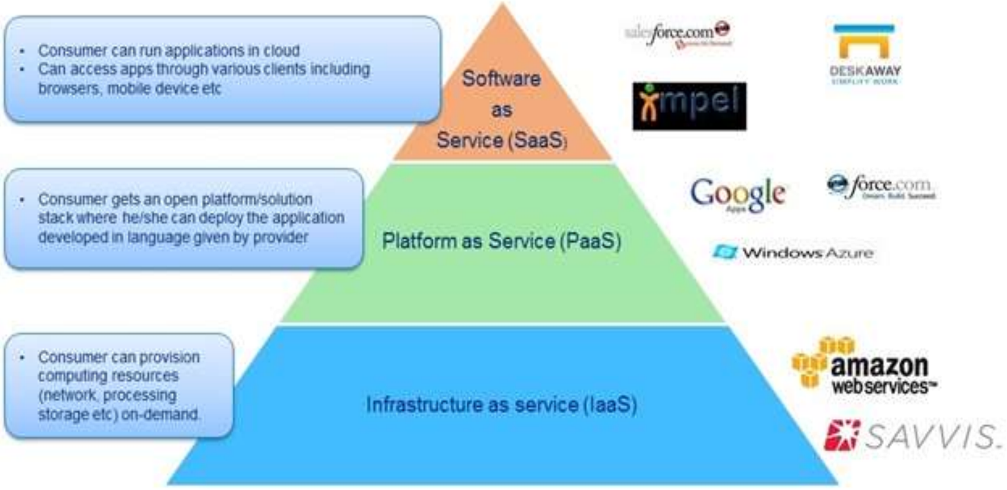
Consumer can run applications in cloud
Can access apps through various clients including browsers, mobile device etc
Software as Service (SaaS)
salesforce.com
impel
DESKAWAY
Consumer gets an open platform/solution stack where he/she can deploy the application developed in language given by provider
Platform as Service (PaaS)
Google Apps
force.com
Windows Azure
Consumer can provision computing resources (network, processing storage etc) on-demand.
Infrastructure as service (IaaS)
amazon webservices
SAVVIS.

**Figure 1.2: Cloud Service Models**

- **Infrastructure-as-a-Service (IaaS)**

This is a first and base layer of the3-level setup. it's accustomed provide any system to interfacing customers and servers and besides offers virtual machines to begin, stop, get to and organize virtual servers and limit squares. The pay-per-use advantage is completed at this layer of the3-level planning. Instances of IaaS square measure Amazon EC2, Windows Azure, Rack space, Google cipher Engine, etc. Infrastructure-as-a-Service like Amazon internet Services offers virtual server event) to begin, stop, get to and mastermind their virtual servers and limit. Within the endeavor, the disseminated process empowers associate association to acquire a similar quantity of limit as is needed, and produce more and more on-line once needed. Since this remuneration for-what-you-use show resembles the means power, fuel and water square measure used, it's sometimes tacit as utility getting ready.

- **Platform-as-a-Service (PaaS)**

This is a second or focus layer of 3-level setup. during this model, a part is given to customers which usually joins an operating system, programming tongues, execution conditions, databases, lines, and internet servers. Delineations square measure AWS Elastic stem, Heroku, Force.com and Google App Engine. Stage as-an advantage within the cloud is delineated as a course of action of programming and issue headway gadgets inspired on the provider's structure. Designers create applications on the provider's part over the net. PaaS suppliers might use Apis, on-line interfaces or section programming bestowed on the customer's laptop. Force.com, (an outgrowth of Salesforce.com) and google applications square measure instances of PaaS. Specialists got to perceive that beginning at currently; there are not any benchmarks for ability or information quality within the cloud. a few suppliers will not enable programming created by their customers to be gotten off the provider's platform.

- **Software-as-a-Service (SaaS)**

This is a third or higher layer of the3-level planning. This model offers "On-demand software" to customers while not foundation arrangement and running of the applications. Customers got to pay and use it through some consumers. Representations square measure Google Apps and Microsoft workplace 365. Within the item as-an advantage cloud illustrates, the dealer provides the gear institution, the item issue and interfaces with the client through a front-end entrance. SaaS is associated particularly wide market. Organizations are something from Web-based email to stock management and info taking care of. Since the professional association has each the appliance and also the information, the top client is permissible to use the organization from where.

## 2. LITERATURE REVIEWS

In a couple of sorts of writing portray the work that officially done in the many papers describedTPA and homomorphic encryption [22], and TPA and cloud. Also, explain About method utilized playout test played out outcomes.

**Cong Wang, Member, IEEE, Sherman S.M. Chow, Qian Wang-Feb 2013, IEEE -**They projected a piece reviews {for completely different|for various} purchasers effectively pack evaluating bolster different records review learning of knowledge. so much reaching security and execution assessment show the projected plans square measure demonstrably secure and passing fruitful. They have empowers associate outer knowledgeable to review client's cloud data while not taking within the data content, completely different allotted evaluating tasks from varied purchasers is performed meantime by the TPA during a security saving manner, macintosh based mostly arrangement has been performed and hashing calculation is employed to perform measuring whereas meantime managing the knowledge. HLA and macintosh based mostly structures have been utilized to play out the wildcat arrangement and contend out the outcomes. the outcomes and execution assessment are tired substitute views, for instance, they need taken some model items and methods the actual outcome tally estimation correspondence organize and sporadic gauze is employed as a touch of this expect to confirm that the TPA wouldn't soak up any data regarding the knowledge substance set away on the cloud server amidst the helpful taking a goose at procedure, that not simply clears out a load of cloud consumer from the dread and maybe expensive auditing task nonetheless likewise decreases the clients' dread of their re-appropriated data spillage. The cloud client (U), United Nations agency has Brobdingnagian live of information reports to be secured within the cloud; the cloud server (CS), that is run by the cloud knowledgeable association (CSP) to present information amassing organization and has vast cupboard space and computation resources (we will not separate cesium and CSP later on); the outcast investigator (TPA), United Nations agency has bent and limits that cloud clients do not have and is trusty to assess the circulated warehousing advantage steady quality for the customer upon inquire.

**Rachna Arora, Anshu Parashar, June-**2013plan for distinctive calculations to wipe out the concerns with relation to data misfortune, isolation, and protection while accessing the web application on cloud. Calculations like: utilized and similar investigation to boot introduced to ensure data on the cloud. calculations, solitary utilized unscrambling while produced within the middle. Planned by, expressly execution obligated things, for instance, inserted framework. AES (Advanced cryptography Standard) was composed by an authority in 2001. Associate open key calculation concocted moreover asymmetric calculation, the calculation that utilizations various decipherment. wide range of calculations square measure distinctive in relevance one another. The key size of AES calculation is the calculation. Calculation creators completely different calculations and considered the outcomes out of the wide range of calculations.

**Fangguo Zhang, Reihaneh Safavi-Naini and Willy Susilo** "An economical Signature theme from linear Pairings and Its Applications" proposed another short mark plot that dissimilar, utilizes capacities, for instance, uncommon capacities. Moreover, the setup matching operations than BLS set up as is more practical than BLS conspire. This mark arranges to develop a hoop mark conspire and another technique for designation [14]. We tend to provide the right security proofs for the new mark plot and also the ring mark conspires within the irregular prophet demonstrate. Another short mark plot that's a lot of productive than BLS conspire. The safety of this mark conspire depends upon another issue, to be specifically incontestable. insight of this basic mark conspire, a hoop mark plot and another strategy for assignment look at, calculation an identical based mostly calculation that is protection safeguarding and prepared to play out {the data|the knowledge|the data} security while not intrusion and concealing the primary information while not interference within the examining procedure help of produces 128 piece enter length keeping in mind the top goal to stay up the knowledge in confirmation, utilized therefore on play out the execution and play out the recreation for the current calculation, the conspire allows a consumer to substantiate that associate underwriter is true. The set up utilizes an identical capability for confirmation and marks square measure amass parts bend. Bend offers protection list scientific discipline assaults against allowing marks. Marks square measure often alluded to marks, marks, just marks. The mark plot set up versatile picked assaults) the exceptive presence of irregular prophets and also the obstinacy of the.

**Zhifeng Xiao and Yang Xiao, IEEE June 2013 conference – "Security and Privacy in Cloud Computing"** They have managed distinctive property order, uprightness, openness, duty, and insurance preservability and played out the diverse security concern issues in viewpoints, makers have methodically analyzed the security and assurance issues in disseminated registering in light of a quality-driven framework, recognized illustrative security/ (e.g., mystery, dependability, obligation, insurance ), and likewise manhandled remembering the ultimate objective to perform distinctive ambushes. Watch methodology and recommendations were discussed likewise, thusly consolidated parts of circulated figuring, the data trustworthiness affirmation made overseeing encryption computation survey estimation open to affirm the regard delivered respectability related archive, managed perspectives, for instance, customer account get the chance to uprightness affirmation framework assurance shielding spill in the midst of.

**Kevin D. Bowers, Ari Juels, and Alina Oprea at RSA Laboratories, Bedford proposed a work titled "Proofs of Retrievability: Theory and Implementation"**They proposed a method a hypothetical structure for the plan of evidence of retrievability, they have demonstrated their work best on the past retrievable systems and they have proposed a variation of the Juels-Kaliski convention and portray a model usage. We show down to earth encoding notwithstanding for records F whose size surpasses that of customer principle memory. additionally, they have chipped away at considering the difficulties experienced when outlining down to earth POR conventions. Initially, we demonstrate to develop external codes that can encode extensive records productively, while as yet safeguarding a high least separation. We characterize and develop pragmatic antagonistic blunder amending codes that, naturally, give no favorable position to an enemy in defiling the encoded document than appropriating defilements arbitrarily crosswise over record squares. Furthermore, as the arbitrary circle gets to is costly, we introduce strategies to encode extensive records incrementally, in just a single go through the document. At long last, they have closed the work that they got yield and accomplishes bring down capacity overhead, endures higher mistake rates, and can be demonstrated secure in a more grounded antagonistic setting. At long last, we gave a Java usage of the encoding calculation of the new variation, in which documents are prepared and encoded incrementally, i.e., as they are perused into fundamental memory and the further work as indicated by them facilitate advancement and well work in sealing of unique record should be possible on the framework.

**Zhuo Hao, Sheng Zhong, Nenghai Yu Proposed A paper** –"A Privacy-Preserving Remote information Integrity Checking Protocol with information Dynamics and Public Verifiability" In this paper, they need to be projected protocol to assist open certainty [10]. The projected protocol supports open certainty while not the help of the associate outsider examiner. Moreover, the projected convention doesn't unleash any non-public information to outsider verifiers. Through a proper investigation, we tend to demonstrate the rightness and security of the convention. From that time onward, through hypothetic examination and check comes regarding and incontestable the execution of their projected work is nice as distinction with the present work a distant data trait checking convention that backings data parts that was projected before in past work, they need taken an endeavor at their work keeping in mind the top goal to method Communication, Computation and Storage prices – the various value produce whereas The correspondences between the champion and also the server happens within the Challenge and GenProof steps, and break down the machine value of consumer, server, and champion .their projected work produces projected convention underpins data flow at the sq. level, which contains piece inclusion, sq. adjustment and piece cancellation. Our convention will while not abundant of a stretch facilitate dynamic data refreshes in lightweight of the very fact that the labeling was relying upon the substance accessible on piece of information, each data tag depends simply on sq., not on another statistics [11] .finally they need to be finished up their work as projected convention bolsters data inclusion, adjustment, and cancellation at the piece level, and moreover underpins open proof [16] . The projected convention is clothed to be secure on associate untrusted server [33]. it's likewise non-public against outsider verifiers, and also the incontestable system by them was concerning cost accounting and tentatively and right down to earth approach was incontestable as best as they resolved to be done and still the work on correct mapping amongst data and labels ought to be increased was aforementioned by their examination future work, the target to accomplish data level.

## 3. PROJECTED WORK

**Proposed work and algorithmic rule Overview:** With a particular finish goal to demonstrate our greatest among the accessible late calculation taken mix is lately cryptography strategy for data security warehousing and more hashing capability procedure SHA-2 is utilizing for the dynamic honesty check method. Our projected framework likewise utilized RSA calculation for the knowledge security and data getting ready for the transfer over the cloud server and recreation performance. The hash work SHA-0 was issued as a government customary by the authority in 1993.SHA-1 distributed because the successor to SHA-0SHA-2published of each 2002 and its variants square measure distributed and at the to wrap things up SHA-224 distributed table three.1 portray here the correlation {of varied|of varied|of assorted} calculations in various parameters.

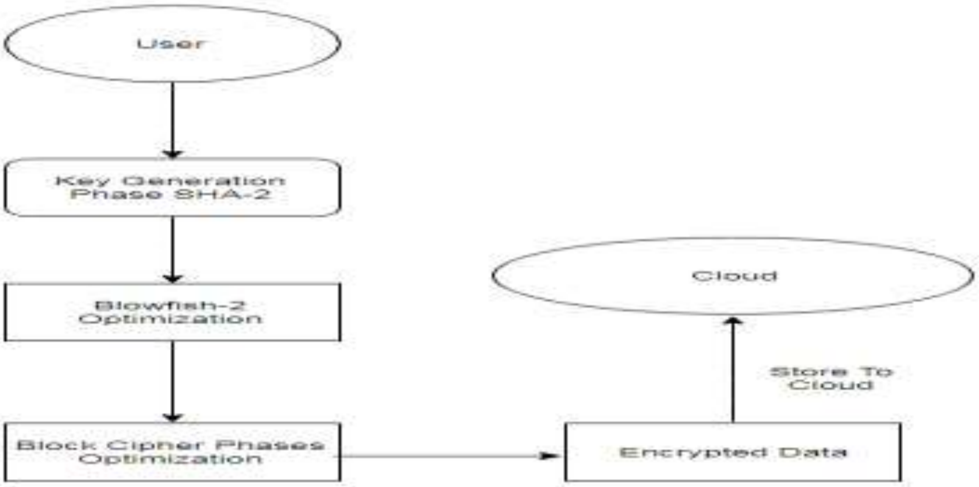

**Figure 3.4: Working Architecture of Proposed Algorithms for Encryption**

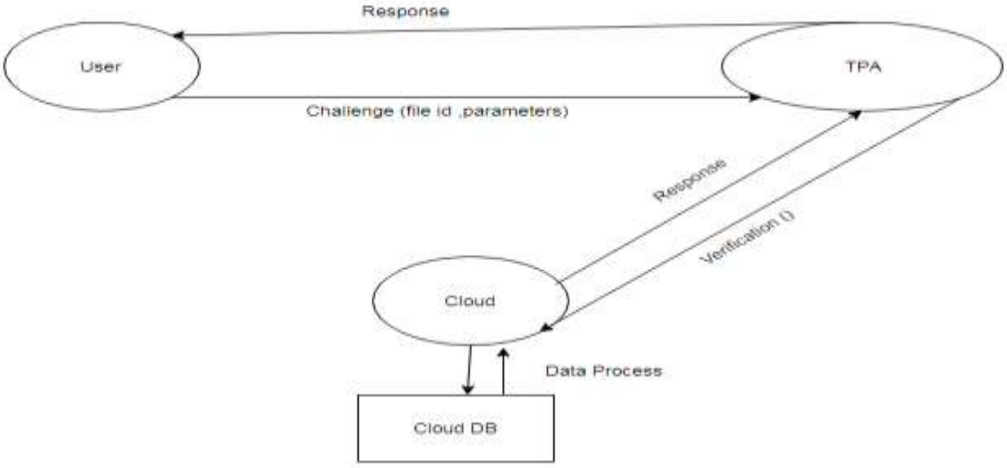

**Figure 3.5: Working Architecture of Proposed Algorithms Verification**

In figure 3.1 and figure 3.2 above the complete working architecture of proposed work and its flow is shown. By understanding, it is being analyzed that how the proposed architecture is working towards.

## 4. SIMULATION AND RESULT ANALYSIS

Reproduction is the bit of equipment or programming that theorizes the conduct of the system without displaying a genuine system. In distributed computing research, cloud reproduction is a method that assesses the conduct of the cloud by figuring the association between different distributed computing gadgets or by numerical equations. In test systems, the cloud is demonstrated with different gadgets, connections, and applications and after that; it is broken down to assess its execution. Clients can likewise redo the cloud test systems to obtain their particular investigation prerequisites. Our work having following necessities for reproduction [12] in the distributed computing:-

i. Cloud Simulator (We Used CloudSim API and Apache Server for Cloud Simulation)
ii. Development Environment (JDK or NetBeans)
iii. Database

### 4.1 Result Analysis

As the requirement of the system and implemented by us here is the comparative analysis is made based on the key size, server computation time, TPA computation time where the system proved our proposed scenario as best among the available technique.

**Table 4.1: Comparison in Terms of Computation Time**

| File Upload Size(in MB) | Computation time (in ms) Existing Algorithm | Computation time (in ms) Proposed Algorithm |
|---|---|---|
| File 1- 250 MB | 359 | 259 |
| File 2- 750 MB | 887 | 525 |
| File 3- 1000 MB | 816 | 509 |

In the table above the data shows the static discussion analysis for computation time in between existing and proposed techniques performed by the system. As per the result monitored above shows the efficiency of our algorithm. This proves the better approach and efficiency of our proposed work system.

**Table 4.2: Comparison in Terms of Cost**

| File Upload Size(in MB) | Computation cost (in INR) Existing algorithm | Computation time (in INR) Proposed algorithm |
|---|---|---|
| File 1- 250 MB | 652 | 252 |
| File 2- 750 MB | 1104 | 987 |
| File 3- 1000 MB | 1430 | 1121 |

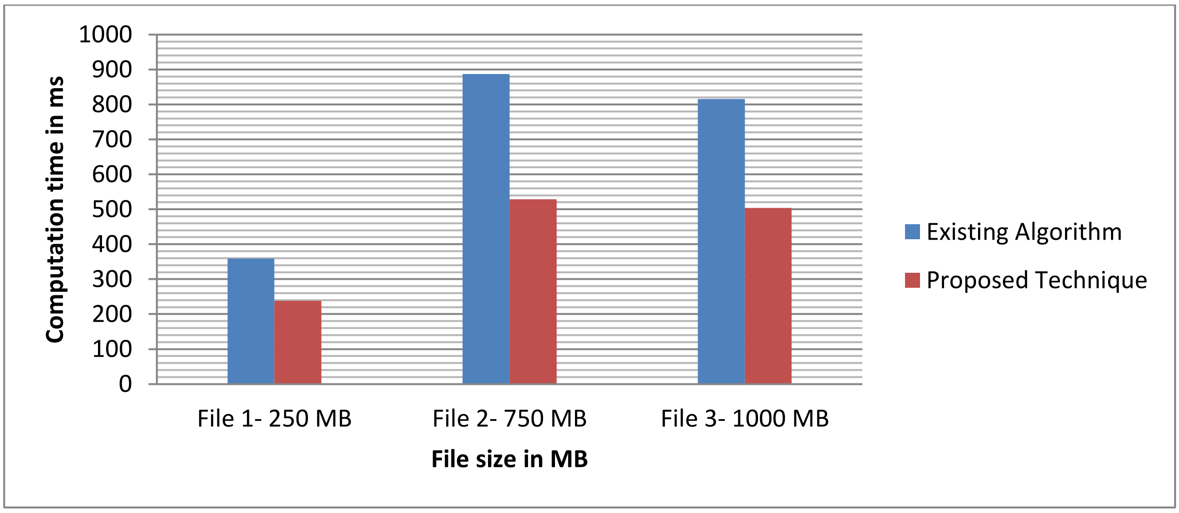

**Figure 4.1: Comparison in Terms of Graph**

In figure 5.4 above the comparison analysis graphically is defined where the system architecture with a given graph is mentioned. In the presence of the graph above it is stated that the proposed algorithm works efficiently while comparing with an existing technique. The proposed algorithm computes low computation time processing the data in a secure manner using Blowfish2 proposed.

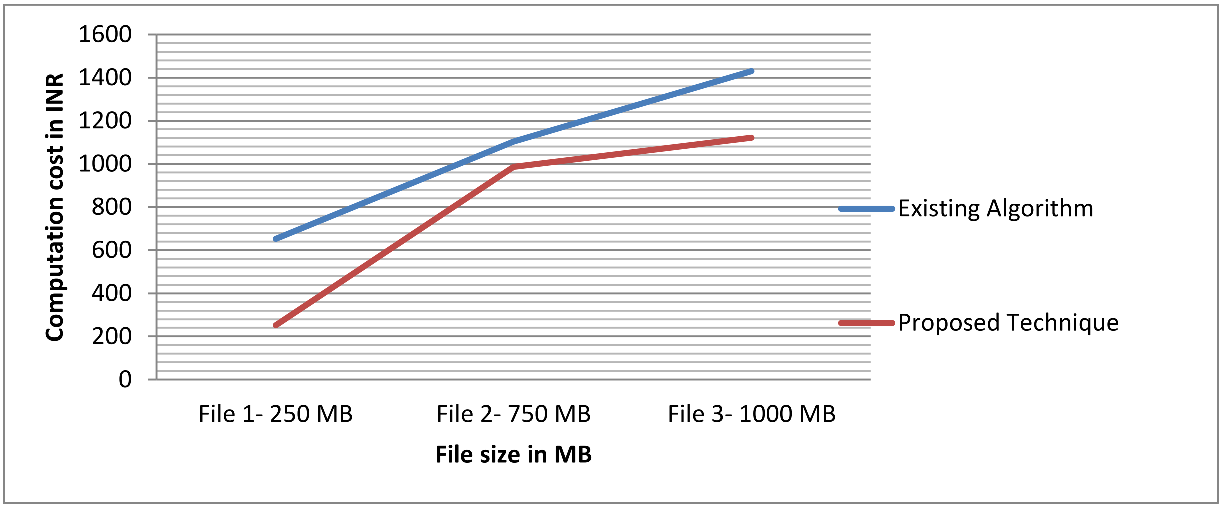

**Figure 4.2: Line Graph of Computation Cost in INR between Existing and Proposed Technique.**

In figure 4.2, line graph of computing, the cost is performed, which shows the comparative analysis of given proposed algorithm. As per the discussed and observed the result of the proposed algorithm and performed an experiment with Apache server framework between the proposed algorithm and existing work. It makes use of enabling privacy-preserving storage and data sharing security in between the cloud component architecture.

## 5. CONCLUSION

Our discourse incorporates into there, for instance, system is unapproved. Moreover, inspects framework being the takes anyway mixed. The considered existing and proposed work was performed and implemented using Java Apache Framework and thus the result was monitored using the different considered parameters. As per our observation, the proposed algorithm is efficient and reliable in terms of encryption scheme which follows symmetric key encryption and also in case of integrity verification which is the latest hashing SHA volume to prove and execute our system over the existing algorithm. Our approach found the advantage of the RSA encryption algorithm over the existing Attribute-based encryption. Which makes the advantage of symmetric key encryption and performed the fast process over the existing work scenario to prove the efficiency of our algorithm? Our work is computed using the parameter computation time which is efficient than the existing attribute-based encryption algorithm.

As per discussion, the proposed algorithm outperforms best in its field where both the encryption and hashing perform best among. Our further work is going to perform a real-time implementation algorithm is computed real scenario application.

## REFERENCES

1) Liu, H., Ning, H., Xiong, Q., & Yang, L. T., “Shared authority based privacy-preserving authentication protocol in cloud computing”, *IEEE Transactions on parallel and distributed systems*, *26*(1), 241-251, 2014.
2) Chen, D., & Zhao, H., “Data security and privacy protection issues in cloud computing”, In *2012 International Conference on Computer Science and Electronics Engineering* (Vol. 1, pp. 647-651). IEEE, 2012.
3) Wang, B., Li, B., & Li, H., ‘’Oruta: Privacy-preserving public auditing for shared data in the cloud. *IEEE transactions on cloud computing*, *2*(1), 43-56. 2014.
4) Los, R., Shackleford, D., & Sullivan.B. ‘’The notorious nine cloud computing top threats in 2013. *Cloud Security Alliance*, 2013.

5) Wang, B., Li, B., & Li, H., ‘’Panda: Public auditing for shared data with efficient user revocation in the cloud. *IEEE Transactions on services computing*, *8*(1), 92-106, 2013.
6) Li, J., Li, Y. K., Chen, X., Lee, P. P., & Lou, W., ‘’A hybrid cloud approach for secure authorized deduplication. *IEEE Transactions on Parallel and Distributed Systems*, *26*(5), 1206-1216, 2014.
7) Pamies-Juarez, L., García-López, P., Sánchez-Artigas, M., & Herrera, B., ‘’towards the design of optimal data redundancy schemes for heterogeneous cloud storage infrastructures. *Computer Networks*, *55*(5), 1100-1113, 2011.
8) Chen, D., & Zhao, H., ‘’Data security and privacy protection issues in cloud computing. In *2012 International Conference on Computer Science and Electronics Engineering* (Vol. 1, pp. 647-651). IEEE, 2012.
9) Ren, K., Wang, C., & Wang, Q., ‘’Security challenges for the public cloud. *IEEE Internet Computing*, *16*(1), 69-73, 2012.
10) Wang, Q., Wang, C., Ren, K., Lou, W., & Li, J., ’’Enabling public auditability and data dynamics for storage security in cloud computing. *IEEE transactions on parallel and distributed systems*, *22*(5), 847-859, 2010.
11) Lou, W., Ren, K., Wang, C., & Wang, Q., ‘’Privacy-Preserving Public Auditing for Storage Security in Cloud Computing. In *Proc. 30th IEEE Int'l Conf. Computer Communications (INFOCOM 10), IEEE Press, San Diego, CA* (pp. 525-533), 2010.
12) Fox, A., Griffith, R., Joseph, A., Katz, R., Konwinski, A., Lee, G. & Stoica, I., ‘’Above the clouds: A berkeley view of cloud computing. *Dept. Electrical Eng. and Comput. Sciences, University of California, Berkeley, Rep. UCB/EECS*, *28*(13), 2009.
13) Zissis, D., & Lekkas, D., ‘’Addressing cloud computing security issues. *Future Generation computer systems*, *28*(3), 583-592, 2012.
14) Zhang, F., Safavi-Naini, R., & Susilo, W., ‘’An efficient signature scheme from bilinear pairings and its applications. In *International Workshop on Public Key Cryptography* (pp. 277-290). Springer, Berlin, Heidelberg, 2004.
15) Zissis, D., & Lekkas, D., ‘’Addressing cloud computing security issues. *Future Generation computer systems*, *28*(3), 583-592, 2012.
16) Wang, B., Li, B., & Li, H., ‘’Oruta: Privacy-preserving public auditing for shared data in the cloud. *IEEE transactions on cloud computing*, *2*(1), 43-56, 2014.
17) MOHAMMED, D. A., & SRAVANTHI, S., ‘’Secure & Data Integrity Proof in Cloud Storage, 2015.
18) Liu, F., Tong, J., Mao, J., Bohn, R., Messina, J., Badger, L., & Leaf, D., ‘’ NIST cloud computing reference architecture. *NIST special publication*, *500*(2011), 1-28, 2011.
19) Bowers, K. D., Juels, A., & Oprea, A., ‘’HAIL: A high-availability and integrity layer for cloud storage. In *Proceedings of the 16th ACM conference on Computer and communications security* (pp. 187-198). ACM, 2009.
20) Erway, C. C., Küpçü, A., Papamanthou, C., & Tamassia, R., ‘’Dynamic provable data possession. *ACM Transactions on Information and System Security (TISSEC)*, *17*(4), 15, 2015.
21) Sengupta, N., & Holmes, J., ‘’Designing of cryptography-based security system for cloud computing. In *2013 International Conference on Cloud & Ubiquitous Computing & Emerging Technologies* (pp. 52-57). IEEE, 2013.
22) Singh, G., ‘’Modified Vigenere Encryption Algorithm and Its Hybrid Implementation with Base64 and AES. In *2013 2nd International Conference on Advanced Computing, Networking and Security* (pp. 232-237). IEEE, 2013.
23) Jaber, A. N., & Zolkipli, M. F. B., ‘’Use of cryptography in cloud computing. In *2013 IEEE International Conference on the control system, computing and Engineering* (pp. 179-184). IEEE, 2013.
24) GR, V., & Reddy, A. R. M., ‘’ An efficient security model in cloud computing based on soft computing techniques. *International Journal of Computer Applications*, *975*, 8887, 2012.
25) Chalse, R., Selokar, A., & Katara, A., ‘’A new technique of data integrity for analysis of the cloud computing security. In *2013 5th International Conference and Computational Intelligence and Communication Networks* (pp. 469-473). IEEE, 2013.

26) Sugumaran, M., Murugan, B. B., & Kamalraj, D., ‘’An Architecture for data security in cloud computing. In *2014 World Congress on Computing and Communication Technologies* (pp. 252-255). IEEE, 2014.

27) Popović, K., & Hocenski, Z., ‘’Cloud computing security issues, and challenges. In *The 33rd International Convention MIPRO* (pp. 344-349). IEEE, 2010.

28) Modi, C., Patel, D., Borisaniya, B., Patel, H., Patel, A., & Rajarajan, M., ‘’A survey of intrusion detection techniques in cloud. *Journal of network and computer applications*, *36*(1), 42-57, 2013

29) Modi, C., Patel, D., Borisaniya, B., Patel, H., Patel, A., & Rajarajan, M., ‘’A survey of intrusion detection techniques in cloud. *Journal of network and computer applications*, *36*(1), 42-57, 2013.

30) Wang, R. Z., Lin, C. F., & Lin, J. C., ‘’Image hiding by optimal LSB substitution and genetic algorithm. *Pattern recognition*, *34*(3), 671-683, 2001.

31) Sahu, C., & Pathre, A., ‘’A cloud data access control and lockbox approach for data sharing. In *2017 International Conference on Information, Communication, Instrumentation, and Control (ICICIC)* (pp. 1-4). IEEE, 2017.

32) Agarwal, A., ‘’Secret key encryption algorithm using genetic algorithm. *International Journal of Advanced Research in Computer Science and Software Engineering*, *2*(4), 216-218, 2012.

33) Du, W., & Zhan, Z., ‘’Building decision tree classifier on private data. In *Proceedings of the IEEE international conference on Privacy, security and data mining-Volume 14* (pp. 1-8). Australian Computer Society, Inc., 2002.