A New Data Security with Privacy-Preserving and Deduplication based Cloud Storage through Public Cloud Auditing in Cloud Computing

Sulthana ASR
Research Scholar
Rathinam college of arts and science
Sulthanaunius07@gmail.com

Dr.K.Juliana Gnanaselvi
Head of IT Department
Rathinam college of arts and science
sunil.juliana@gmail.com

Abstract: The field of Cloud computing is an thought-provoking arena of research that enables user accumulating and preserving data on cloud servers and offers inexpensive, measurable, simultaneously efficient and reliable outsourcing storage services. Currently surrendering of private data to cloud service providers (CSPs) is accomplished by various person and companies in terms of the benefits of cloud storage over expenses and maintenance. Anyways, this innovative data storage structure still meets several new security and efficiency threats. This research proposes Confidentiality and Trust Aware-Preserving Deduplication Cloud Storage Scheme promoting public cloud auditing (CTAPDA) with that motivation. Comprises of four steps, the research work at first proposing an effective recovery technique for retrieving the deleted data based on Hybrid Flower Pollination and Gaussian function based Genetic Algorithm (HFPGFGA). In second step, safer file deduplication on encrypted file that helps reliable implementation atmosphere for deduplication of cloud storage structure for the particular copy, also identifies secure verification tag deduplication. In third step, The CTAPDA program uses the Dual Integrity Convergent Encryption technique to guarantee the privacy of data throughout the process of file deduplication and integrity audit. At final, apart from helping every data owner to initiate their own files' integrity auditing by themselves, the proposed scheme also helps regularly in consignment of cloud server to the unknown auditor to perform several auditing responsibilities simultaneously for safeguarding the outsourced records integrity. Also, the reliability which is the significant factor achieved by the CTAPDA scheme's. Also simulation experiments and numerical analysis are carried out to validate its performance.

Keywords: Cloud Storage, Public Cloud Auditing, Batch Verification, Trusted Execution Environment, Encryption, Secure Deduplication.

1. Introduction

Cloud Computing is becoming more prominent and is gaining a raising exposure in the science and commercial communities. Cloud Computing is considered as top most relevant innovations revealed by Gartner in his research and with good prospects for organizations and companies in successive years. Cloud computing provides omnipresent, simple, challenging network access to wide range of customizable computing resources such as storage, servers, services, software, and networks, that can be easily distributed and delivered with minimal maintenance or service provider involvement. Cloud computing is one which possess efficient delivery architecture and its major intention is to come up with Safe, Fast, Versatile data storage and network computing with all computing resources envisioned as services and distributed through internet.

The cloud stimulates collaboration, mobility, scalability, flexibility, the ability to adapt to demand-specific volatility, speeds up development process and creates cost cutting opportunities through automated and efficient computing. This Cloud Computing technology offers various challenges apart from various benefits. Adoption of Cloud Computing has several features, meanwhile some notable hurdles are also there. Security is one of the most important obstacles to adoption, followed by enforcement concerns, privacy and legal issues [2].

Since cloud computing is a relatively recent computing paradigm, there is a lot of debate about how protection can be accomplished at certain stages and how protection of applications is relocated to Cloud Computing. The instability has frequently driven executives of data to claim the safety is their great anxiety with cloud computing. There are several security risk based on areas like "general" internet dependency, inability to control, multi-tenancy and internal security integration and external information storage. The cloud has several certain characteristics compared to conventional technologies, such as its enormous size. The fact that resources assigned to cloud vendors are entirely distributed, heterogeneous and completely virtualised. Traditional safety structures such as identification, authentication, and authorisation are no longer adequate in their present form for clouds [4].

Nevertheless, because of the cloud service structures utilized, the operating structures and the technologies used to facilitate cloud services, cloud computing may provide an entity with various risks than conventional IT solutions. Regrettably, incorporating protection into such solutions is still recognized as more static[5]. Transferring important applications and

private information to common cloud surface is of significant anxiety to those companies that expand beyond their regulated data center network. To address these issues, a cloud solution operator should confirm that consumers remain to get the same security protection and privacy safeguards on their applications and services, furnish consumers with documentation that their entity is safe and can fulfil their service-level agreements, and demonstrate compliance to auditors.

We are trying to propose a privacy and confidential implementation environment-conserving deduplication cloud storage with general cloud auditing through this research, based on the current outstanding work. A novel Confidentiality And Trust Aware-Preserving Deduplication Cloud Storage scheme supporting public cloud auditing (CTAPDA) is proposed, which concurrently helps reliable deduplication of files on encrypt files and deduplication of validation tags. Single Copy along with distinct set of validation tags in the cloud for all file is maintained by CSP. This strategy also offers communal reliability confirmation of the specific file contained in cloud deduplication by using prevailing cloud auditing technologies. Our efforts can be summed up more accurately, as follows:

1) Throughout this study, a HFPGFGA (Hybrid Flower Pollination and Gaussian function based Genetic Algorithm) based multi-server architecture is addressed, for restoring missing data using multi-cloud backup servers.
2) The TEE for managing the key and for serving third party servers is accomplished using stable deduplication stratagem
3) The CTAPDA scheme utilizes the convergent encryption technique for ensuring the generation of same ciphertext from same plaintext file ,thereby realizing encrypted files deduplication. Therefore, during the stage of deduplication and public audit, the CSP and TPA can not access the original data of the DO's file and the secrecy of the outsourced data is assured.
4) This CTAPDA strategy not only encourages of DO in autonomously delegating the own files integrity audit to the TPA along with assisting CSP for simultaneous auditing tasks administration which in turn assures the outsourced files reliability.
5) Submit a strategy to get stable deduplication as well. The safe encryption of their data is semantically accomplished using the Dual Integrity Convergent Encryption method and the server can verify the user's permission to access a file on the basis of user privileges.
6) Numerical analysis and simulation experiments include the efficiency comparisons between the CTAPDA schemes and the contemporary schemes.

The structure of the paper is specified in various sections: Section 2 provides the outcomes of structured analysis. In Section 3 explains the proposed Confidentiality And Trust Aware-Preserving Deduplication Cloud Storage scheme supporting public cloud auditing (CTAPDA). At final, section 4 gives the exploratory results with discussion and some findings and section 5 deal with conclusions with future research.

2. Related Work

This segment summarizes recent work on the legality of storage of Cloud data .Some of the strategies that use soft client applications without the need for additional hardware are discussed. Kamara at el. [6] anticipated prototype of a full safe storage system without stating anything about implementing the involved components. In recent time, few researchers have suggested third party auditor (TPA) based strategies. Wang on el. [7] recommended a method that enables Cloud data storage protection to be audited publicly through external TPA deprived of requiring data local copy or placing additional accessible burdens on Cloud. This developed a publicly auditable data protocol, which keeping up privacy of data but Xu et al.[8] find this scheme to be highly flawed as it is weak to prevent tag falsification. Sookhakat el. [9] adopted a methodology in which the information is secure with the aid of public auditing. In this method, the encryption is utilized for the purpose of storing the information with the advantages of any accessibility loss and any functionality for authoritative users loss.

Wang et al. (2011) [10] improved PoR "proof of retrievability" by elucidating the standard Merkle Hash Tree (MHT) structure proficient for validating block tags. It is an expanded form of their initial scheme[7] lacking in proof of protection, both of these approaches completely assist dynamic operations of the information and privacy is not maintained. Zhang and Blanton (2012)[11] used a data structure to tackle the problem of data integrity termed as balanced update tree. The frequency of updates plays a vital role in the tree size which is carried out for remote data blocks and entire quantity of outsourced data is completely autonomous. Every data dynamic operation is preceded by validation of the server's performance accuracy, their scheme removes such verification process.

Ni et al. (2013) [12] recommended a technique for dynamic privacy auditing with the help of physical attack. Wang et al. (2010b)[7] suggested that editing of cloud data can be accomplished by any illegal user identified in the course of the audit. Indeed, alternative solution is delivered with respect to the original protocol version properties. Luca Ferretti et al.[13] suggested a combination of cloud storage infrastructure and data security for execution of simultaneous encrypted cloud databases operations. Also it mitigates the restriction of elasticity,

accessibility and scalability of built-in assets in cloud-computing results. Hong Liu et al.[14] projected a methodology for determining cloud storage security concerns by utilizing authority-based privacy sustaining authentication protocol particularly suited for multi-user cloud applications. KanYag et al. [15] recommended a policy named Cipher text-Policy Attribute-based Encryption (CP-ABE) for cloud storage data access accomplishment by effective data access control scheme and revocable for cloud storage systems with multi-authority. The defined approach accomplished safety both forward and backward.
In [16] proposed a stratagem for steady and operative query range by establishing data perturbation namely random space perturbation (RASP) and for securing cloud data, kNN query services is utilized. The features such as encryption, dimension expansion, random noise injection and random projection to provide high resistance to data and query attacks. It also retains multi-dimensional ranges, enabling for the implementation of current indexing strategies to accelerate range query processing. The kNN- queries are processed by utilizing kNN algorithm with the RASP query algorithm.

In [17], local-recoding for big data anonymisation in contrast to violations of proximity confidentiality are considered for finding functional solution. In particulara proximity privacy structure is defined for semantic proximity of sensitive values and multiple sensitive attributes. The local recoding as a proximity-conscious clustering issue is also analysed. In this flexible two-phase aggregation method is suggested for clustering of t-ancestors and a proximity-conscious agglomerative clustering algorithm.

In [18] an active and reliable dynamic audit protocol is greatly involved in reassuring data holders for proper storage of data. In this cloud storage systems audit process is initiated by an efficient and privacy-conserving audit protocol. Audit protocol is also utilized for facilitation of information dynamic processes which is basis of effective and fundamentally secure random oracle framework.

In [19], recommended a new secrecy-conserving system that facilitates pooled data public auditing, which is warehoused in cloud. Also ring signatures are involved in quantifying the authentication metadata for pooled data accuracy audit. The reliability of pooled information deprived of recovering widespread data is attained by public validators which in turn accomplished by securing signer in-block of pooled data description. With our system, the of the signer in-block of shared data is kept private by public validators, who can effectively validate the reliability of pooled information deprived of recovering widespread data. This scheme was in essence an ABE based data-sharing system. Since most of the above schemes are built on the basis of bilinear pairing, their computational complexity is considerable. Therefore an efficient data-sharing system based on encryption method is proposed with honesty in this research.

3. Proposed Methodology

There would be enormous quantity of data at the data center, since the cloud customer can able to keep numerous data into cloud storage. The data may get eradicated from the data centers by man-made catastrophe (either customer itself, without their knowledge or CSP) or by natural disasters (either volcanoes or earthquakes). Data has been produced in large quantities which need the services or techniques of data recovery. Consequently, the design of an effective data recovery method is required to restore the missing data. Some researchers have suggested various strategies for data recovery but lack accuracy and efficiency. The PoW protocol in proposed CTAPDA plays key role in restriction of confines cloud users with insufficient privileges from retrieving privileged data where the Hybrid Flower Pollination and Gaussian function based Genetic Algorithm (HFPGFGA) is being used to backup the data. Every user is delegated with a set of advantages in this method, during the system configuration process. A privilege-based authentication code is generated if a user attempts to upload a file to the cloud. CSP decides based on authentication code whether the user has the permission to implement a duplicate search or a PoW operation; the file can only be retrieved by the user who proves that is their own file. This sort of system considerably provides more immune to the brute-force guessing and Man-in-the-Middle attack. This CTAPDA scheme implements a trustworthy cloud storage structure that enables the user to safely store by encoding the client side data and decoding the data after downloading from the cloud using Dual Integrity Convergent Encryption. The proposed CTAPDA architectural diagram is shown in Fig.1.

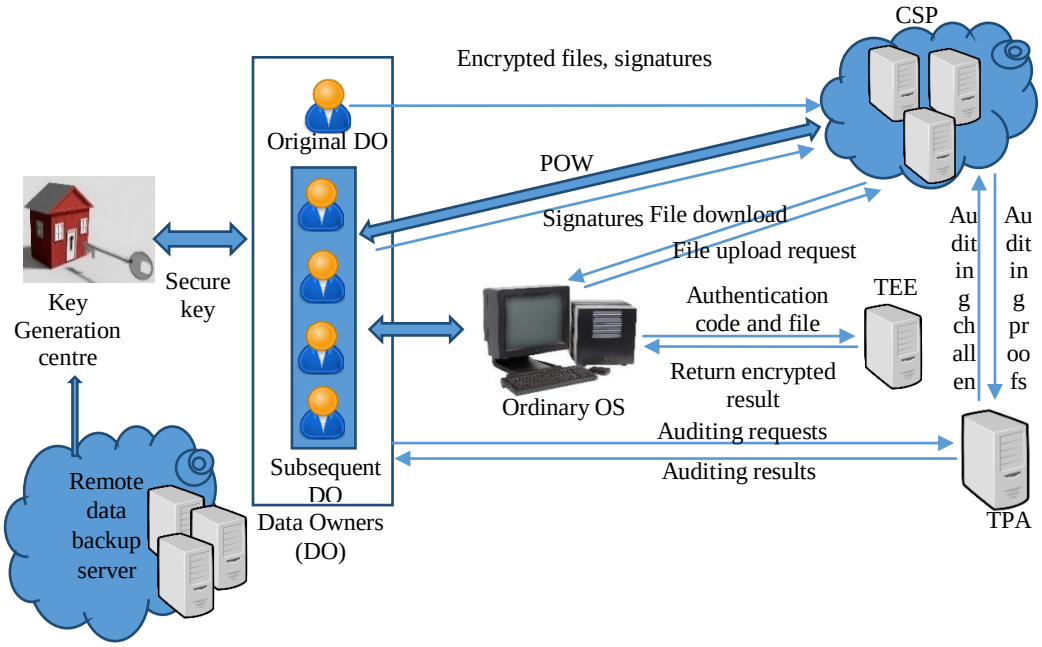

Fig.1.The proposed architecture diagram of CTAPDA

3.1. System Model

In general, there are four key entities involved in a file deduplication method, that facilitate public cloud auditing:

Key Generation Center (KGC): The major entity called KGC, which initializes the process parameters and creates a portion of the public keys.

Data Owner (DO): Data Owner is an entity, which has a huge amount of data to maintain in the cloud servers, may either be an person or an association. Because, various DOs may have the similar file in the cloud, the certain file DOs are separated into the actual DO (only one) and the subsequent DOs (maybe more than one) as per the order in which the file is uploaded.

Cloud Service Provider (CSP): It is nothing but an entity with considerable space to store and computing resources to offer the DO with storage for data and safe deduplication services.

Third Party Auditor (TPA): An entity trusted by DOs has proficiency and the ability to validate the ethics of the outsourced file.

Trusted Execution Environment (TEE): TEE's special key-management system is the foremost solution for customer-side encryption, as the scheme is in full compliance with encryption. We introduce our scheme to display that it increases protection and entails reduce overhead costs compared to earlier deduplication methods.

As demonstrated in Fig. 1, at first, picking of seed is accomplished by the KGC and Dos are given right direction by means of secure channels, followed by outsourced file and the seed plays a vital role in generating the convergent key. Meanwhile, DOs encrypt the outsourced file and create the respective block signatures. Unless the CSP does not possess the same file, the original DO will upload the file's ciphertexts and their CSP signatures; or else following file DOs will execute POW protocol together with the CSP. If the later DOs transfer the POW protocol, a direct link is granted to the file cloud storage and their own signatures are uploaded to the CSP. When the primary cloud system lacks its data and cannot provide users with data, the proposed strategy enables the customer with the ability to obtain information from any backup system. In addition, the TPA must implement an integrity audit with the CSP on requests from DOs for auditing and finally, the auditing evaluations to DOs would be retorted by TPA. Through a 'proof-of-ownership' (PoW) protocol, TEE-based scheme prohibits access to privileged data by cloud users with inadequate advantages.

3.2. Proposed Data Recovery Technique Using Hybrid Flower Pollination and Gaussian Function based Genetic Algorithm (HFPGFGA)

There is a risk of data loss and it may also leads to economic crisis, if the system crashed or any form of natural or human catastrophe occurred in consistent operation of a business. The initial authentic data can be restored by using some of the data recovery techniques. A technique is required to achieve efficiency and reliability in order to restore the missing original data, since the current recovery strategies are not reliable and efficient. The reproduced copies of data are retained in more than one database for data recovery. If data loss happens at one place, it can be recovered from another backup server using HFPGFGA. This subsequent steps of the process is as follows:

1. Initialization - Cloud Initialization.
2. Evaluation - Total file in cloud (i.e)Computing the size and status of the cloud service provider, where the flower pollination algorithm is applied with its efficient execution along with exploration and exploitation output.
3. Selection - Selection of user.
4. Crossover - Comparing data files in cloud storage with Gaussian crossover for the end user [20].
5. Mutation - Erased/Deleted file gets retrieved.
6. Termination - Restoration is completed.

Flower pollination algorithm

The key stages of the generic Flower Pollination Algorithm (FPA) are emphasized as follows.

Step 1: The algorithm begins by defining the original values of the most critical variables including population, size pop, probability of switching and the highest number of files in the cloud

Step 2: The primary population $x_i, i = \{1,\dots,n\}$is created at random and the fitness function of each solution $f(x_i)$ in the population$p$ is assessed by computing the appropriate objective function

$$TOTCost = Cost * size\ of\ file\ in\ KB$$

Step 3: Repeat the following steps until the termination criteria is met, which is to meet the required number of generations in the backup server.

Step 3.1: The universal pollination cycle begins by creating a random number $r$for every solutionsx_i, where $r\in$ [0, 1]

Step 3.2: If r < p, The recent solution is created as follows across a Lévy distribution.

$$X_i^{t+1} = X_i^{t+1} + \mathfrak{L}(X_i^t - g_{best}^*)$$

Where g_{best}^* is the best solution currently available,$\mathfrak{L}$ is a Lévy flight,$\mathfrak{L} > 0$ and determined by the formula below

$$\mathfrak{L} \sim \frac{\lambda\Gamma(\lambda)\sin\left(\frac{\pi\lambda}{2}\right)}{\pi} \times \frac{1}{S^{1+\lambda}}, s \gg s_0 > 0$$

$\Gamma(\lambda)$is the standard gamma function and this distribution is valid for large steps $s > 0$.

Step 3.3: Or else, the domestic pollination procedure begins by creating a random numberϵ, $\epsilon \in [0, 1]$ and the new solution is created as follows

$$X_i^{t+1} = X_i^t + \in \left(X_j^t - X_k^t\right)$$

Where X_j^t, $X_k^t$are pollens (solutions) from the various flowers of the same genus. If X_i^t said to possess local random walk if it originates from the same genus or picked from the same population.

Step 3.4: Assess every solution in the population and upgrade population solutions as per their target values

Step 3.4: Rate the solutions and take the best solution currently availableg_{best}^*

Step 4: Deliver the best solution identified until now.

Procedure of HFPGFGA based Uploading and Recovering

The basic system of the proposed HFPGFGA algorithm is set out in Algorithm 2 and the key phases of the suggested algorithm is specified below.

Step 1: The proposed HFPGFGA algorithm begins by defining its variable values including the probability of switching p,the variables number in each divisionν, the amount of solutions in each population sizeη, the number of partitions Part-no, the Gaussian crossover probability$P$, geneticmutation probability P_m and highest number of iterations $\max_{itr}$.The user uploading file $F$is the users measured as the population to the N cloud servers.

Step 2: The counter $t$iteration is initialized and the original population is created spontaneously From file F, hash code H1 is created and placed in database.

Step 3: For each solution By applying one of the two Flower Pollination operators (either Global Pollination or Local Pollination), a new solution X_i^{t+1} is created for each solution X_i^t, in the population, relying on the figure of the random number generated $r$and the probability of switching p.

Step 4: Assess the fitness feature of all population solutions and choose a middle population from the existing one. The fitness task here is to measure the file size from which the TOT cost and new balance are calculated.

TOT Cost=Cost * file size in KB, New Balance=Available Balance – TOTfiCost. Also check the New Balance < 0 condition, If it meets the criteria then, the process will end; or else, upload the file to the database and revise the balance.

Step 5: In order to maximize the complexity of the quest and to solve the dimensional issue, the existing population is divided into ν × η, where ν is the number of files at each division while η is the number of solutions at each population. Here, the user must also choose the file, which needs to be downloaded.

Step 6: To prevent premature convergence, the genetic mutation operator is implemented in the whole population. If the file is removed then the backup server could restore it.

Step 7: The solutions in the population are assessed by measuring its fitness function, increasing the iteration

counter and repeating the overall operations until it met with the termination requirements. Pick the file to be downloaded and create the key code H2.

Step 8: Ultimately, we furnish the best possible solution. When both hash codes are the same, the original file and the authenticity of the file will be verified.

3.3. Secure Deduplication and Public Cloud Auditing

In particular, the Setup phase basically deals with system parameters and creating keys using Dual Integrated Convergent Encryption Process with Decryption [21] and it is demonstrated in Fig.2

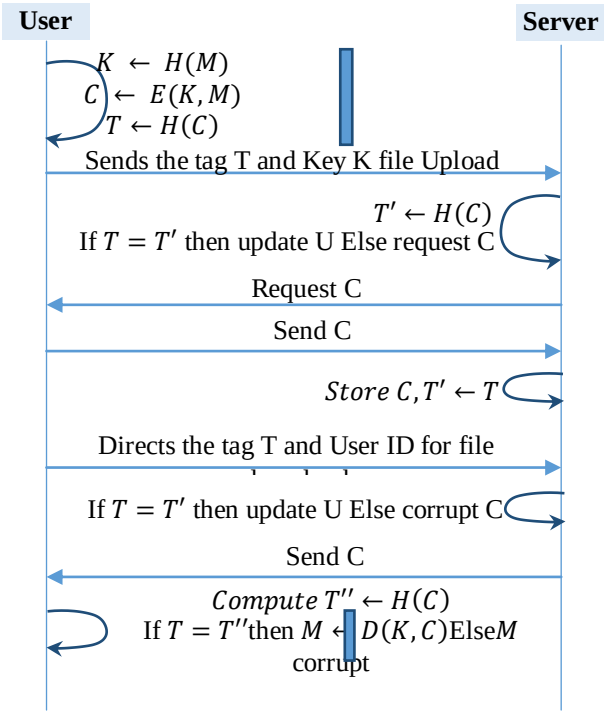

Fig.2.The Dual Integrated Convergent Encryption (DICE) Process with Decryption

The client only holds K and the tag T all through the overall Dual Integrated Convergent Encryption process, while the server keeps T ' and C. This technique transfers just the tag T compared with current encryption methods and thus the bandwidth requirement is minimised.

The ***Storage Phase*** is nothing but the data file deduplication phase involving three cases:

Case 1 :Initial DO of the file which stores the original ciphertext file;

Case 2 :The following DOs act together with the CSP for file deduplication storage space.

Case 3 :The outsourced file can be accessed by the user, if and only if he/she has the appropriate files and credentials, a file-based authorization code and the privilege are needed when the user demands the file.

The ***Auditing Phase*** is intended for file integrity audits of the specific copy in the cloud deduplication storage system, permitting each DO to autonomously examine the integrity of its own file at any time.

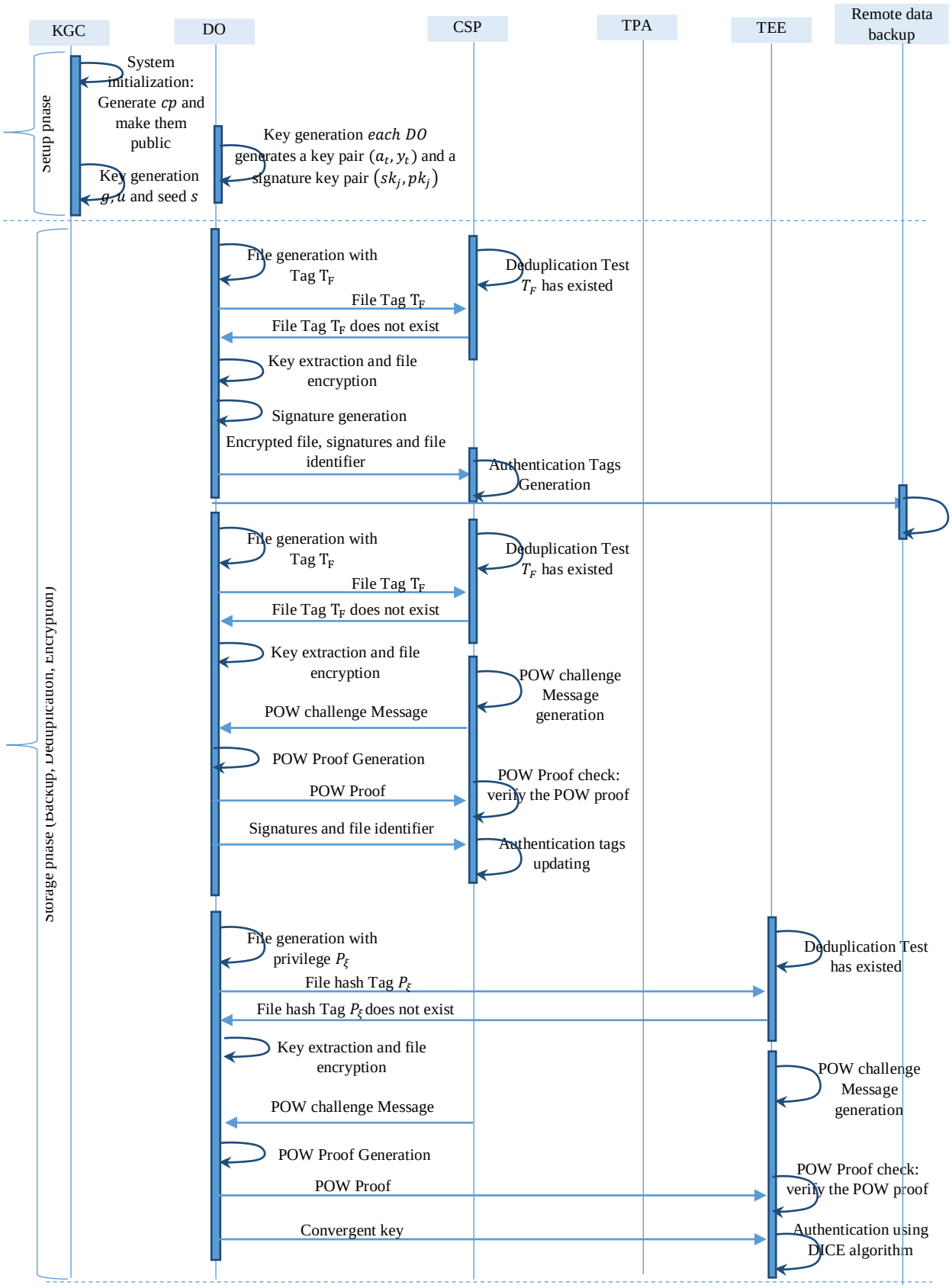

Fig.3.Setup and Storage phase of CTAPDA

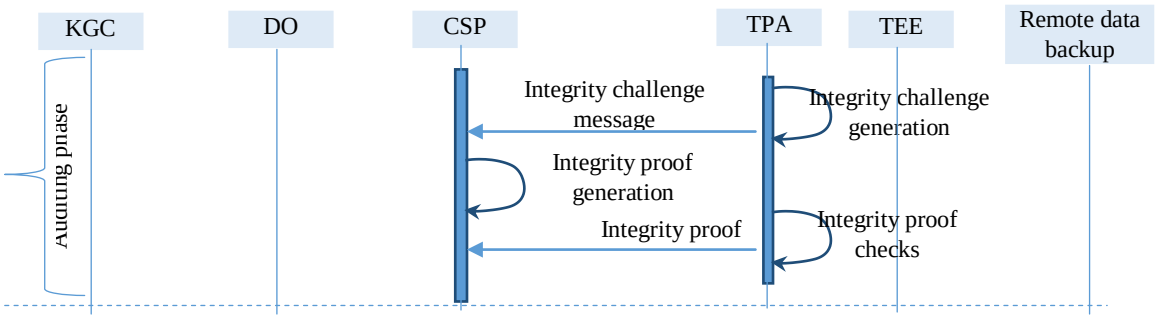

Fig.4.Auditing Phase of CTAPDA

Setup Phase: This phase includes of the two steps as follows:

Step 1 (System Initialization): The Key Generating Centre (KGC) creates ciphertext for the system parameters(cp)$cp = (p, G, G_T, e, H, Hash, Enc, Dec, Extractor, (n,k)RSC)$ and sorts it to be public, where p is a large prime which involves security parameter 1^{λ}, G and G_T are multiplicative cyclic groups with order p, e is a bilinear pairing $e: G \times G \rightarrow G_T$, H and Hash are two collision-resistant secure hash function with $H: \{0,1\}^* \rightarrow G$, (Enc, Dec) is a pair of symmetric encryption and decryption algorithms such as DICE, Extractor is one utilized for key extraction that is computed by the content of the file, and $(n,k)RSC$ is a Reed-Solomon code.

Step 2 (Key Generation): Initiall two random elements $g, u \in G$y are selected by the KGC and create them public. Then, Seed is randomly selected by the KGC and on the sly directs it to authentic DOs. A random value is chosen by each data owner $DO_t \forall (t = 1,2,\cdots,W) a_t \in Z_p^*$ and sets $y_t \leftarrow g^{a_t}$, and produces a pair of secret and public keys (sk_t, pk_t) for signature, and then every DO creates(y_t, pk_t) public and retains (a_t, sk_t) secret. Storage Phase: To attain together confident client side deduplication and integrity auditing, the storage phase of the CTAPDA scheme is dissimilar from those of classical auditing systems. In detail, our CTAPDA scheme has three different file uploading procedures for the original owner and subsequent owners.

Case 1: File uploading for the original owner of file $F$when the file F is not stored in the CSP, the first uploader of file F is deliberated as the original owner DO0. Thus, the file uploading practise for the original owner DO0 is specified in the subsequent.

• Step 1 (File Tag Generation): The DO_0makes the file tag $T_F = Hash(F)$ and transfers T_F to the Cloud Service Provider (CSP).

• Step 2 (Deduplication Test): The CSP runs the deduplication test by examination whether the file tag $T_F$now prevails in the CSP's local storage. If $T_F$does not exist, the DO_0 and CSP remain to accomplish step3-5.

• Step 3 (Key Extraction and File Encryption and Split): Firstly, the DO_0compute encryption $key\ K = Extractor(F; s)$. Secondly, the DO_0 encrypts the file F with the symmetric key K to get $C_F = Enc(F, K)$. Finally, the DO_0applies the Reed-Solomon code $(n,k)RSC$ on C_F to obtain $C_F' = \{C_1, C_2, \cdots, C_n\}$ such that CF can be entirelyimproved from any k blocks among $\{C_1, C_2, \cdots, C_n\}$.The fault tolerance of the deduplication cloud storage system is achieved using Reed-solomon codes to encode CF.

• Step 4 (Signature Generation): The DO_0calculates the signature for each block with$\sigma_i^0 = H(T_F \| i) \cdot u^{C_i}$, $i = 1,2, \ldots, n$. Furthermore, the DO_0 computes the file identifier $ID_F^0 = T_F \| n \| sign_{sk_0}(T_F \| n)$, where $sign_{sk_0}(T_F \| n)$ is used to make sure the integrity of the file tag T_F . Then, the DO_0 uploads the file set $C_F' = \{C_1, C_2, \cdots, C_n\}$, the signatures set $\Phi_0 = \{\sigma_1^0, \sigma_2^0, \ldots, \sigma_n^0\}$ and the file identifier ID_F^0 to the CSP.

• Step 5 (Authentication Tags Generation): For each encrypted block C_i, the CSP further generates an authentication tag θ_i as $\theta_i = \theta_i^0 = e(\sigma_i^0, y_0), i = 1,2, \cdots, n$ based on the signatures$\Phi_0 = \{\sigma_1^0, \sigma_2^0, \ldots, \sigma_n^0\}$. At last, the CSP should store the authentication tags $\Lambda = \{\theta_1, \theta_2, \ldots, \theta_n\}$.

Case 2: The consequent owners of file F is uploaded. When a DO possess a file F and need to store in CSP, the evidence should be revealed for specific file possession F. Legal subsequent owner of file F is considered and denoted by DOt if POW is valid. In consequence, the file uploading system for a subsequent owner DOt of file F is as follows.

• Step 1 (File Tag Generation): The DO_t generates the file tag $T_F = Hash(F)$ and sends $T_F$to the CSP.

• Step 2 (Deduplication Test): The CSP runs the deduplication test. The ownership proof of CSP protocol with the DO_t is verified if there exist duplication of file.

• Step 3 (Key Extraction and File Encryption and Split): Firstly, the DO_t compute encryption key $K = Extractor(F; s)$ based his file F. Secondly, the DO_t encrypts the file F as $C_F = Enc(F, K)$. Finally, the DO_tapplies the Reed-Solomon code $(n,k)RSC$ on CF to obtain C_F' such that C_F can be completely recovered from any k blocks among C_F'.

• Step 4 (Proof of Ownership (POW)): Random selection of a d-element subset $D = \{s_1, s_2, \cdots, s_d\}$ from set $[1, n]$ and a set of coefficients $\{v_i\}_{i\in D}$is accomplished where $v_i \in Z_p^*$ and then, he sends the challenge message $chal\_M = \{i, v_i\}_{i\in D} \in D$ to the O_t . For each block C_i, the DO_t generates a signature σ^t with the public key u and his secret key a_t, which is labelled as $\sigma_i^t = (H(T_F||i) \cdot u^{c_i})^{a_t}, i = 1,2, \ldots, n.$ Once acceptance is obtained, the $chal\_M$ from the CSP, theDO_t would yield an ownership proof, comprising the signature proof and the data proof. The signature proof σ^t is fundamentally the signature aggregation of the challenged blocks, i.e., $\sigma^t = \prod_{i\in D}(\sigma_i^t)^{v_i}$. And the data proof is essentially the linear combination of sampled blocks, i.e., $\mu = \sum_{i\in D} v_i C_i$. Finally, the DO_tresponds the CSP with $P_{POW} = (\sigma^t, \mu)$. The proof P_{POW} is significant in computing the CSP

$X = \prod_{i\in D} H(T_F||i)^{v_i} \cdot u^{\mu}$ (1)

and authenticates the ownership proof by examining the following (2):

$e(\sigma^t, g) \cdot e\left(X, \prod_{j=0}^{t-1} y_j\right) = e(X \cdot y_t) \cdot \prod_{i\in D} \theta_i^{v_i}$, (2)

where$\{\theta_i\}_{i\in[1,n]}$ are the authentication tags stored in the CSP. If X holds, it yields Accept and an access link to the file F will be established for the O_t ; otherwise, Reject.

• Step 5 (Signature Uploading): When the proof of ownership are accepted, the DO_tuploads the signature set$\Phi_t = \{\sigma_1^t, \sigma_2^t, \ldots, \sigma_n^t\}$ and the file identifier $ID_F^t = T_F \| n \| sign_{sk_t}(T_F \| n)$to the CSP, and secretly protects the dual integrated convergent key K and the access link in his local storage. Subsequently, the DO_tdeletes the file F and 8t from his local storage.

• Step 6 (Authentication Tags Updating): Later receiving the signature set Φ_t from the DO_t, the CSP computes $\theta_t^i = e(\sigma_i^t, g)$ and then updates the authentication tags $\theta_i$as $\theta_i \leftarrow \theta_i \cdot \theta_i^t$, $i = 1, 2, \cdots, n.$ In the end, the CSP stores the updated tags $\bigwedge$.

Case 3(trusted execution environment): A privilege based authentication code with the assistance of TEE is utilized if a uploading of file to the cloud happens.

When a DO_trequests for uploading content to the remote data center, DO_t first directs the file and the encryption privilege set P_ξ from CA to TEE.Assume that a privilege set $P = \{p_1, p_2, \ldots, p_n\}$, $n \in N^*$ is predefined. Simultaneously, each user is assigned with a privilege subset P_ξ , $P_\xi \subseteq P$ and $m = |P_\xi|$. The symmetric key k_ξ for each privilege $p_i \subseteq P_\xi$ is derived from TEE by a shared protocol.After receiving the file and P_ξ, TEE computes a hash tag $\varphi_F = H(F)$based on the file $F$and generates the privilege key k_ξ according to the received privilege set as $k_\xi = \left(P_1 \| \ldots \| p_i \| \ldots \| p_m, m_{d_m}\right)$ for $p_i \in P_\xi$, $m_{d_m}$is a 'dummy message' that is predefined corresponding to $P_m$for achieving secure encryption.. Beside with these two values, an authentication code $\alpha_F = HMAC_{k_\xi}(\varphi_F)$ is calculated and directed back to the user. Upon receipt of α_F, the DO_tthen directs it to the cloud. Once the CSP gets the authentication code, it first checks whether the corresponding file exists in its storage. In this context, we consider the authentication code as an access key for data deduplication. If the CSP does not possess the authentication code α_F, the user is realized as an initial uploader. Otherwise, $DO_t$is reflected as a subsequent uploader.

Auditing Phase: In a deduplication cloud storage system, the CSP stores only a unique copy of the same file. Storage hardware or software failure might destruct the integrity of a unique file copy. Thus, the DOs need to regulate the integrity of their outsourced file in any time. Usually, limited by the computation power or constrained resources, the DOs may entrust a TPA to conduct separate integrity auditing for file F. Suppose a DO_tdelegate a TPA to do the following steps. For clarity, assume there are $W + 1$ (the number can freely and dynamically increase) DOs for the file F.

• Step 1 (File Identifier Check): To verify the integrity of F, the TPA first validates the validity of signature $sign_{sk_t}(T_F \| n)$. If the substantiation fails, the TPA quits by outputting False; otherwise, it authorizes the file tag T_F and the total number n of blocks for file F, and goes to Step 2.

• Step 2 (Integrity verification): The TPA arbitrarily picks a c-element subset $I = \{s_1, s_2, \cdots, s_c\}$ from set $[1, n]$ and a set of coefficients $\{w_i\}_{i\in I}$, where $w_i \in Z_p^*$. Then calculates a random masking based on the DO_t's public key as $R \leftarrow y_t^r$, where $r \in Z_p^*$is a random number, and sends the challenge message $chal\_M = (\{i, w_i\}_{i\in I}, R)$ to the CSP.

After receiving $chal\_M$, the CSP would yield a data integrityresponse, encompassing of tag proof Θand the data proof. The tag proof 2 is fundamentally the aggregated authenticator of authentication tags of the challenged blocks, i.e.,

$$\Theta = \prod_{i\in I} \theta_i^{w_i}$$

Then, the CSP calculates the linear combination of sampled blocks and aggregates all DOs' public keys except the DO_t's public key as trails:

$$C = \sum_{i\in I} w_i C_i, Y = \prod_{j\in[0,W]\setminus I} y_j$$

The data proofs are composed of $\omega_1$and ω_2, where $\omega_1 = e(u, R)^C, \omega_2 = e(u, Y)^C$. Finally, the CSP sends $P = (\Theta, \omega_1, \omega_2)$to the TPA as the integrity proof.

3) Integrity Proof Check: Later receiving the integrity proof P, the TPA first computes the aggregate value of public keys except the DO_t's public key:

$$Y = \prod_{j \in [0,W] \setminus I} y_j$$

Further, the integrity verification is accomplished by by checking

$$\omega_1 \cdot \omega_2^r \cdot e\left(\prod_{i \in I} H(T_F||i)^{w_i}, R \cdot Y^r\right) = \Theta^{\mathrm{r}}$$

If Θ^{r} holds, it outputs Ture; otherwise, outputs False.

3.4. Security Analyses of CTAPDA

This subdivision outlines the security model and examines the security of CTAPDA scheme. Usually, the TPA is considered to be truthful, indicating that the TPA does the integrity audit process but there is no certainty about the outsourced file. Some time, a unreliable integrity auditing may be caused by CSP which in turn deceive the TPA by means of fake audit facts. In addition, limited file may be acquainted with the limited user by malevolent user and pretends to be the file owner. Consequently, POW and integrity audit plays a significant role in security model of CTAPDA. The CTAPDA system acts as CSP in integrity auditing process. Secondly security concept in [9] is a key role in CTAPDA so that (k, θ) cannot be cheated in the POW process.

CSP-unforgeable: A CTAPDA acts as CSP- if no deceptive CSP may effectively manipulate the auditing evidence deceiving the TPA with an insignificant possibility.

(k, θ)-uncheatable: A CTAPDA system is said to be (k, θ)-uncheatable if specified a file F with min-entropy k, no harmful user, who knows incomplete file containing θ-bit Shannon entropy of F, convinces the CSP of having a non-negligible possibility of all file F.

Against poison attack: The above statements is due cause for the conflict prevention feature of the hash functions, it is computationally difficult for the opponent to conjecture the plain text M in such a way that both conditions, $T = T'$ and $C = C^*$ are true.

Safety of encryption key: The various types of keys are extracted to fulfil several encryption algorithms. The encryption keys are: privilege key, convergent key. We presume that such a situation occurs where a hardware adversary dedicates himself to dispossessing all confidential data. When these keys are stored in the local computer, it would seem the attacker controls these keys. TEE is perceived as a black box from outside the universe, illegitimate attackers cannot learn the TEE keys. While, it is also prohibited to look at the internal implementation of TEE, which means what kind of encryption algorithm we have selected is hidden to outside attackers. And even though the attacker contains the details contained in TEE, illegally purloin of any information is not possible by the attacker.

PoW checking protocol accuracy: The evidence value in the PoW checking protocol is created and validate the accuracy by randomly picking the user privileges and the hash value of the file retained. Notably, every time there is always different evidence value for PoW protocol performance. Concurrently, The XOR function is used with hash values and security parameter ensuring that no original privilege set and file hash value is seized for security factor particulars. Thereby collapsing occurs if any forging of reversing the pow protocol happens.

4. Experimental Results and Discussions

Within this section, numerical analysis and computational experiments are performed with respect to factors such as communication, computing and storage costs. The detailed performance comparisons between CTAPDA and the prevailing related schemes PCAD[22] and P-PCAD[23], both of which are deduplicated by the public cloud audit is obtained. The comparison reveals that original file size F is almost same in all schemes and the cost factor for communication ,uploading F and storage of F are not considered .

4.1. Communication Cost Comparison

First, we will compare the CTAPDA strategy with the current PCAD and P-PCAD strategy regarding the communication charges in the storage and audit process, and the results will be shown in Fig. 3. The user in PCAD scheme [12] need not submit the authentication tags, due to the integrity of file F audit with the support of authentication tags uploaded through the original possessor. Conversely, the original possessor entails high communication resources in the storage procedure and could not assure the integrity of file F audit cannot be accomplished by every possessor exclusively. The P-PCAD arrangement will make sure that each user examines the authenticity of file F individually. The CSP maintains entire authentication tags from complete file F users ensuing in high overhead and storage costs. The PCAD at [13], the value of audit phase communication is minimal. The PCAD on encrypted file does not endorse secure deduplication. From the results it is clear that the CTAPDA scheme is much more effective in terms of communication expenses in the course of the storage and audit phases. The explanation is that it not only helps each DO in individually delegating the integrity audit of their own files to the TPA, nonetheless encourages CSP in occasionally delegating TPA to deal with manifold auditing responsibilities simultaneously safeguarding the integrity of outsourced files with minimum bandwidth consumption.

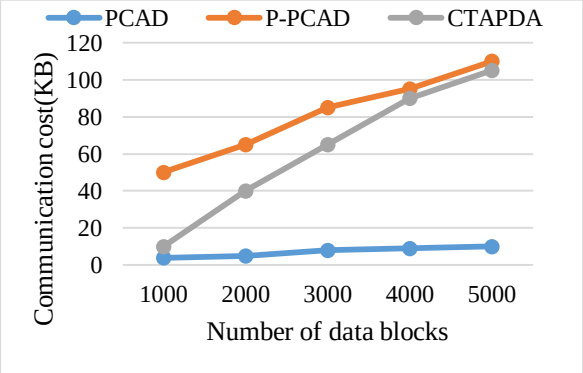

Fig.5.Comparisons of communication costs

4.2. Computation Cost

Now, present them in Fig. 6, the CTAPDA scheme with current schemes such as PCAD and P-PCAD. The computational complication only determines the multiplication process, exponential operation and bilinear pairing technique on a multiplicative group to explain computational values clearly and effectively.

The main drawback of PCAD scheme is that high computation storage resource compared to P-PCAD. The CTAPDA system decreases the computation expenses of creating the data owner's authentication tags, since partial measurement is accomplished by high computational resources. The suggested CTAPDA structure requires the minimum computing resources in the storage process and the evidence is generated. To summarise, we may infer that our CTAPDA scheme is fairly successful in computational aspects during the storage and audit phase.

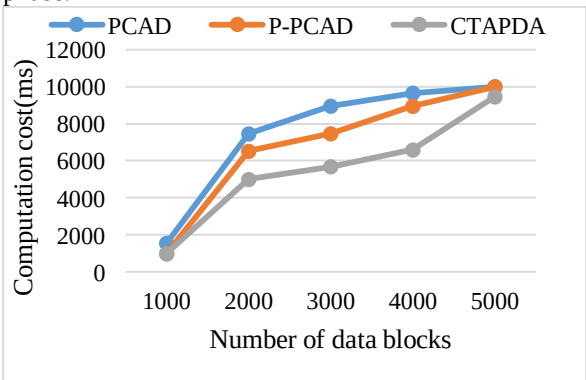

Fig.6.Computation Cost Comparison

4.3. Execution Time

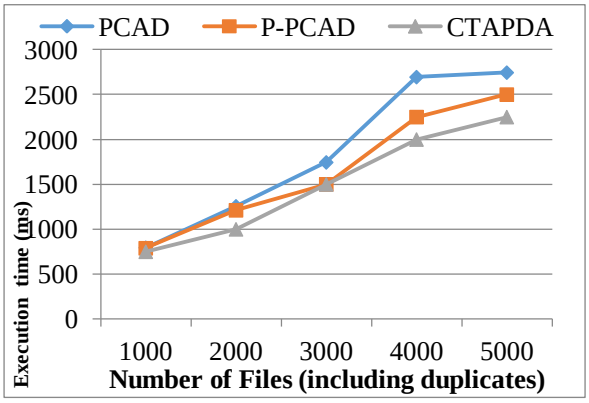

Fig.5. Energy consumption Vs No. of Nodes

Fig.5 indicates the link between implementation time communications and file numbers. The proposed CTAPDA method could be said to take lower time compared to the other PCAD and P-PCAD method. From the figure, the time value will decrease considerably, if malware files launch cloud attacks. CTAPDA will diminish the time value of implementation compared to PCAD and P-PCAD as it is insusceptible to the poison attack. Also when compared with convergent encryption system, the mode reveals higher data deduction.

5. Conclusion and Future Work

The research work concentrates on privacy-maintaining cloud storing of information backup and client-side deduplication facilitates public cloud auditing and CTAPDA. Specifically, the CTAPDA system recognizes deduplication using Dual Integrity Convergent Encryption technique to construct the similar ciphertext from the same plaintext file. All together, we turn down the communication bandwidth and the time needed for the computation. The basic concept driving our approach is to carry out integrity tests on both ends. Particularly, the CTAPDA scheme can recognize secure deduplication of authentication tags from several Dos via combining authorization tags from the identical blocks. Furthermore, in the course of deduplication , convergent encryption and random masking secure data privacy is achieved and integrity auditing with the aid of CTAPDA scheme.

In specific, the CTAPDA scheme benefits every DO in independently delegating the integrity audit of its own records to the TPA, however assists CSP in delegating TPA for outsourced files batch auditing regularly. Additionally, the suggested technique based on swarm is being used for remote data backup for data restoration. At final, we systematically demonstrated the protection of the proposed system. Statistical survey and experimental results designated that the proposed system is effective when compared with prevailing stratagems with respect to computation and storage costs. We plan to expand this protocol in the future by checking it for other

vulnerabilities beyond the toxic attack, such as dictionary attack, recognition, and side channel.

References

1. Gartner Inc: *Gartner identifies the Top 10 strategic technologies for 2011*. Online. Available: . Accessed: 5-Jul-2011 http://www.gartner.com/it/page.jsp?id=1454221 Online. Available: . Accessed: 15-Jul-2011
2. Khalid A: Cloud Computing: applying issues in Small Business. *International Conference on Signal Acquisition and Processing (ICSAP'10)* 2010, 278–281.
3. KPMG: From hype to future: KPMG's 2010 Cloud Computing survey.. 2010. Available: http://www.techrepublic.com/whitepapers/from-hype-to-future-kpmgs-2010-cloud-computing-survey/2384291 Available:
4. Rosado DG, Gómez R, Mellado D, Fernández-Medina E: Security analysis in the migration to cloud environments. *Future Internet* 2012, 4(2):469–487.
5. Cloud Security Alliance: *Security guidance for critical areas of focus in Cloud Computing V3.0.*. 2011. Available: https://cloudsecurityalliance.org/guidance/csaguide.v3.0
6. Kamara, S., Lauter, K.: Cryptographic Cloud Storage. In: Proceedings of the 14th International Conference on Financial Cryptograpy and Data Security, FC 2010, pp. 136– 149. Springer, Heidelberg (2010)
7. Wang, C., Wang, Q., Ren, K., Lou, W.: Privacy-preserving public auditing for data storage security in cloud computing. In: 2010 Proceedings IEEE INFOCOM, vol. 54(2), pp. 1–9 (2010)
8. Xu, Chunxiang, Xiaohu He, and Daniel Abraha-Weldemariam. "Cryptanalysis of Wang's auditing protocol for data storage security in cloud computing." In *International Conference on Information Computing and Applications*, pp. 422-428. Springer, Berlin, Heidelberg, 2012.
9. Sookhak, Mehdi, Abdullah Gani, Hamid Talebian, Adnan Akhunzada, Samee U. Khan, RajkumarBuyya, and Albert Y. Zomaya. "Remote data auditing in cloud computing environments: a survey, taxonomy, and open issues." *ACM Computing Surveys (CSUR)* 47, no. 4 (2015): 1-34.
10. Wang, Q., Wang, C., Ren, K., Lou, W., Li, J., 2011. Enabling public auditability and data dynamics for storage security in cloud computing. IEEE Trans. Parallel Distrib. Syst. 22 (5), 847–859
11. Zhang, Y., Blanton, M., 2012. Efficient dynamic provable possession of remote data via update trees. IACR Cryptol. ePrint Arch., 291.
12. Ni, J., Yu, Y., Mu, Y., Xia, Q., 2013. On the security of an efficient dynamic auditing protocol in cloud storage. IEEE Trans. Parallel Distrib. Syst. 25 (10), 2760 1761.
13. Luca Ferreti, Michele Colajanni, MircoMarchetti, "Distributed, Concurrent and Independent Access to Encrypted Cloud Databases", IEEE Transactions in Parallel and Distributed Systems, Vol. 25, No. 2, pp. 437-446, Feb 2014.
14. Hing Liu, HuanshengNing, QingxuXiong, Laurence T. Yang, "Shared Authority Bassed Privacy-preserving Authentication Protocol in Cloud Computing", IEEE Transactions in Parallel and Distributed Systems, Vol: pp:99, 2014.
15. Kan Yang, XiaohuaJia, "Expressive, Efficient and Revocable Data Access Control for Multi-Authority Cloud Storage", IEEE Transactions in Parallel and Distributed Systems, Vol. 25, No. 7, pp1735-1745, July 2014.
16. H. Xu, S. Guo and K. Chen, "Building Confidential and Efficient Query Services in the Cloud with RASP Data Perturbation," in IEEE Transactions on Knowledge and Data Engineering, vol. 26, no. 2, pp. 322-335, Feb. 2014.
17. X. Zhang et al., "Proximity-Aware Local-Recoding Anonymization with MapReduce for Scalable Big Data Privacy Preservation in Cloud," in IEEE Transactions on Computers, vol. 64, no. 8, pp. 2293-2307, 1 Aug. 2015.
18. K. Yang and X. Jia, "An Efficient and Secure Dynamic Auditing Protocol for Data Storage in Cloud Computing," in IEEE Transactions on Parallel and Distributed Systems, vol. 24, no. 9, pp. 1717-1726, Sept. 2013.
19. B. Wang, B. Li and H. Li, "Oruta: privacy-preserving public auditing for shared data in the cloud," in IEEE Transactions on Cloud Computing, vol. 2, no. 1, pp. 43-56, Jan.-March 2014.
20. Kubicki, M., &Figurowski, D. (2018, May). An introduction to a novel crossover operator for real-value encoded genetic algorithm: Gaussian crossover operator. In *2018 International Interdisciplinary PhD Workshop (IIPhDW)* (pp. 85-90). IEEE.
21. Agarwala, A., Singh, P., &Atrey, P. K. (2017, October). DICE: A dual integrity

convergent encryption protocol for client side secure data deduplication. In *2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC)* (pp. 2176-2181). IEEE.

22. J. Yuan and S. Yu, ‘‘Secure and constant cost public cloud storage auditing with deduplication,’’ in Proc. IEEE CNS, National Harbor, MD, USA, Oct. 2013, pp. 145–153,
23. C. Li and Z. Liu, ‘‘A secure privacy-preserving cloud auditing scheme with data deduplication,’’ Int. J. Netw. Secur., vol. 21, no. 2, pp. 199–210, Mar. 2019,