Design of an Improved Data Deduplication Technique for Cloud Storage

Sunil Gupta[1*], Rajeev Bedi[2], Amandeep Kaur[3]

[1,2,3]Beant College of engineering and technology

[*]*Corresponding Author: skgbcetgsp@gmail.com*

Abstract- Exchanging data over the system is generally utilized quick and solid hotspot for correspondence. Clients from wide devotion utilize this component for exchanging and getting to data. Portability and between operability inside cloud framework through disconnected and online mediums are persistently alluring yet the issue of security emerges amid the transmission process. Security and unwavering quality is the key issue during the exchange process which is considered in this exploration. Data security is given utilizing the public and private key block level de-duplication. The analysis is inferred at disconnected information as well as an online information, for example, googledocs. Redundancy handling mechanism of the component is utilized to guarantee that space at information storage supplier is slightest utilized since taken a cost in DSP is went with a measure of capacity utilized. Overall space necessity if there should arise an occurrence of heavy documents is decreased and security of online data getting to is improved by the utilization of Byte level de-duplication.

Keywords: de-duplication, redundancy handling mechanism, cryptosystem

I. INTRODUCTION

In recent year the fast development of media transmission and web, data security turns out to be increasingly huge. Cryptography is the most ideal path for ensuring mystery data. Cryptosystems can be separated into two kinds, secret key cryptosystem, and public-key cryptosystem. The main sort (Secret key cryptosystem), utilizes a similar encryption key to encipher the plaintext and decode the figure content.For this reason, this kind is additionally called as a symmetric cryptosystem. In spite of the fact that secret key cryptosystem is effortlessly to actualize because of less calculation, it has a few disadvantages, excessively numerous keys, key conveyance issue, verification, and nonrepudiation issue. The imperative kind which is the public-key cryptosystem is created to take care of the issues of a symmetric cryptosystem, and RSA cryptosystem is the most prevalent approach. As of late, information security has turned into an essential issue for public, private and guard associations as a result of the extensive misfortunes of illicit information get to. To shield secret information or data from unapproved access, illegal changes and propagation, different sorts of cryptographic strategies are utilized. One of these critical strategies is cryptography which is the investigation of writing in secret frame and it is isolated into two sorts: symmetric and asymmetric cryptography.

Cloud computing becomes the need of the hour nowadays but many experts argue about it [4]. Highly scalable services are provided by the cloud. Users can utilize the services on a pay per use basis. Cloud computing theoretically provides infinite resources but due to the growing number of users, practically services and resources becomes limited. The services and resources required to be distinguished on the basis of the scale of utilization along with cost. Although energy consumption and starvation problems nowadays, associated with cloud computing but still improvement in services could lead to the better framework for concurrent users to access resources more than capable of the machine user hold and hence leads to more popularity and user community attracted towards the cloud.[5].

Cloud interoperability is required during the transmission of data to and from the cloud servers. The cloud service provider ensures QoS(quality of service) through security mechanisms. The security mechanisms used may or may not use redundancy handling mechanism to conserve space. In the proposed system security mechanism along with redundancy handling mechanism is enforced for ensuring the quality of service. The attributes considered for evaluation are described as under.

1.1 Attributes
Before some of the attributes will be defined, the term cloud should be explained. Cloud computing used widely from a long time and provides an opaque framework where services are visible to the user but internal working is hidden[6]. Key attributes in cloud computing are described in this section:

- **Service-Based***:* Cloud main objective is to provide a service-oriented framework by hiding details and showing only the necessary features to the user. This mechanism is also termed as an abstraction.
- **Scalable and Elastic***:* services associated with the cloud are not fixed. Services and be added as and when required depending upon mass usage of services. In other words, the scalable environment is provided by cloud computing[7]. Elasticity in the framework indicates resources are provided on different platforms accessible by multiple users at a time. In other words, concurrency is supported through the use of cloud computing framework.
- **Shared:**[8]Pool of resources are provided by the use of advanced computing environment. Resource if free can be accessed by any number of resources provided resource is not exclusive in nature. The exclusive resource cannot be shared and that resource accessing required queue to be maintained.
- **Metered by Use**: multiple payment modes are supported by cloud infrastructure[9]. Services are accessed on a pay per use basis. Service provider and clients are bound by the service level agreement. The user need to pay for accessing the services mentioned within SLA. The problem however is, even if service is down for the period of time, still user is required to pay for that service.
- **Uses Internet Technologies**: Services are delivered to the user through the use of the internet. Protocol such as hypertext transfer protocol(HTTP), file transfer protocol(FTP), Terminal network(Telnet) etc. are used for this purpose[10].

Symmetric calculations are commonly viewed as quick and they are appropriate for preparing an expansive stream of information. A portion of the popular and proficient symmetric calculations incorporate Two fish, Serpent, AES, Blowfish and IDEA. Also, there are non-specific calculations which offer an elective system for encryption. When all is said in done, hereditary calculations contain three essential administrators: multiplication, hybrid, and change. Then again, there are distinctive well known and productive deviated calculations including RSA, NTRU, and Elliptic curve cryptography.

II. RELATED WORK

[11] Describes the IBE technique with outsourcing computation and also offloads the key generation operations to Key Update Cloud service provider. It also focuses on the critical issues of identity revocation. It accomplishes consistent productivity for both calculation at PKG and private key size at the client, User needs not to contact with PKG amid key-update, as it were, PKG is permitted to be disconnected subsequent to sending the disavowal rundown to KU-CSP, No protected channel or client verification is required amid key-update amongst client and KU-CSP.

[12]proposed the main MCL-PKE scheme without blending operations and gave its formal security. Our mCL-PKE takes care of the key escrow issue and disavowal issue. Utilizing the mCL-PKE conspire as a key building block, it proposed an enhanced way to deal with safely share sensitive information out in the public clouds. This approach supports quick denial and guarantees the classification of the information put away in an untrusted open cloud while authorizing the entrance control strategies of the information proprietor. The exploratory outcomes demonstrate the productivity of fundamental mCL-PKE scheme and enhanced approach for people in general cloud. Further, for various clients fulfilling similar access control arrangements, the enhanced approach performs just a solitary encryption of every datum thing and lessens the general overhead at the information owner.

[13]Proposed a variation of CP-ABE to effectively share the various hierarchical documents in distributed computing. The hierarchical documents are scrambled with an incorporated access structure and the ciphertext parts identified with characteristics could be shared by the records. Thus both ciphertext storage and time cost of encryption is saved. The proposed system has benefits that clients can decode all approval documents by figuring secret key once. Therefore, the time cost of decryption is also saved if the client needs to decode various documents. Additionally, the proposed plot is ended up being secure under DBDH suspicion.

[14]design a virtual encryption card framework that gives encryption card usefulness in virtual machines. In this framework, it displayed the vEC-PPM, which deals with the encryption resource plan. It saved clients' information utilizing a trusted equipment of virtualization in view of TPM. It additionally settled a trusted chain amongst clients and encryption cards in light of the composed protocols. The design of the virtual encryption card empowers the security and productivity of the encryption benefit. A usage examination shows that the effectiveness of the framework is similar to that of the native mode. Later on, it proceeds with the examination, trying to plan a virtual encryption cards bunch to help higher encryption speed and more reasonable similarity with virtualization.

[15]proposed a safe billing protocol for smart applications in distributed computing. It utilized homomorphic encryption by adjusting the Domingo-Ferrer's plan, which can perform different number arithmetic operations to fulfill smart grid billing necessities in a safe way. This plan keeps up the exchange off amongst security and versatility contrasted and other homomorphic plans that depend on either secure, yet inelastic in terms of arithmetic operations assortment. Additionally, it proposed an instrument that guarantees both security and integrity during correspondence between substances. The execution of the proposed system is very satisfactory; it is sufficiently productive to use in lightweight

applications and can be helpfully connected to cloud-based applications.

[16]propose a CP-ABE scheme that gives outsourcing key-issuing, decryption and keyword search work. This scheme is productive since it just needs to download the fractional decryptionciphertext relating to a particular keyword. In this scheme, the tedious matching operation can be outsourced to the cloud specialist organization, while the slight operations should be possible by clients. In this way, the calculation cost of the two clients and trusted specialist sides is limited. Besides, the proposed plot supports the capacity of keywords look which can enormously enhance correspondence effectiveness and further ensure the security and protection of clients. It is difficult to stretch out given KSF-OABE plan to help get to the structure represented by the tree in. [17]In this paper, based on contingent intermediary communicate re-encryption technology, an encrypted information sharing plan for secure distributed storage is proposed. The plan not just accomplishes communicate information sharing by exploiting communicate encryption, yet in addition accomplishes dynamic sharing that enables adding a client to and expelling a client from sharing gatherings dynamically without the need to change encryption open keys. Besides, by utilizing intermediary re-encryption innovation, this scheme empowers the intermediary (cloud server) to specifically share encoded information to the objective clients without the intercession of information owner while keeping information security, so significantly enhances the sharing execution. In the meantime, the rightness and the security are demonstrated; the execution is broke down, and the test comes about are appeared to confirm the possibility and the productivity of the proposed plot.

[18]proposed diagram encryption scheme just makes utilization of lightweight cryptographic natives, for example, pseudo-arbitrary capacity and symmetric-key encryption, instead of moderate homomorphic encryptions. Accordingly, the proposed graph encryption scheme is well disposed to a wide arrangement of graph information based distributed computing and edge registering applications, for example, interpersonal organizations, e-maps, criminal investigations, and so on. Contrast with graphanonymization comes nearer from database group, the proposed system achieves higher security level as the chart itself is encoded and it doesn't make any suspicions on the sorts of attacks.

[19]discussedsecurity enhancement mechanisms including a symmetric, public key and homomorphic cryptosystems to enable experts to comprehend encryption plans for information on distributed storage. AES is utilized as a part of most secure applications for information on distributed storage. Completely homomorphic encryption plans are promising for cloud condition, however, a long way from being useful due to their execution rate. Homomorphic assessment of AES has fascinating applications as a

reasonable encryption conspire for information on distributed storage.

[20]proposed an Improved Encryption Calculation (EEA) for securing the data in cloud stockpiling. This is a symmetric encryption calculation. It utilizes the same key for encoding and unscrambling the data previously put away into the cloud.[21]proposed a lightweight accessible open key encryption (LSPE) conspire with semantic security for CWSNs. LSPE decreases countless calculation escalated operations that are received in past works; along these lines, LSPE has sought execution near that of some useful accessible symmetric encryption schemes.[22] proposed a protected cloud data encryption framework, named the Circulated Ecological Key (DENK in short), with which all records are encoded by one encryption key got from numerous coordinating keys which are keys gotten from approved clients' secret key keys and a believed PC's natural key.[23] proposed to present an effective and unquestionable FHE in light of another mathematic structure that is without commotion.

[24]described various way which is used in cloud computing for data security. The information is put away on to an incorporated area called data centers having a substantial size of information storage. In this way, the customers need to put stock in the supplier on the accessibility and additionally information security. Before moving information into general society cloud, issues of security gauges and similarity must be tended to. A trusted screen introduced at the cloud server that can screen or review the operations of the cloud server. In limiting potential security trust issues and additionally sticking to administration issues confronting Cloud computing, an essential control measure is to guarantee that a solid Cloud computing Service Level Agreement (SLA) is set up and kept up when managing outsourced cloud service suppliers and particular cloud merchants. Cloud computing guarantees to change the financial matters of the server farm, yet before sensing and managed information move.

To resolve the problem with the existing literature proposed literature present efficient solution. The encryption mechanism with the redundancy handling mechanism is proposed as described in the next section.

III. METHODOLOGY

The methodology of the proposed work consists of the registration process at first place. The registration in the proposed system will be a two-phase process. In the first phase, registration at data storage provider is made. After successfully registering, the user can load files at data storage provider end. To generate keys users require performing registration at the key service provider. In order to retrieve the files, users must log in to the DSP and then KSP. The keys generated could be used in order to decrypt the file. The mechanism also uses the redundancy handling mechanism

for preserving space for extra file loading. Also, an online source of files like Googledocs can be used to retrieve the files and perform encryption and decryption.

The detailed steps are described as under
1.1 Registration at DSP

The registration at DSP comprises of unique username and password. Username and password once registered at DSP can be used for accessing file uploading module
1.2 Registration at KSP
The key service provider(KSP) is used in order to generate the keys for the file which is uploaded. The proposed system is capable of generating the keys for files generated from the online source.
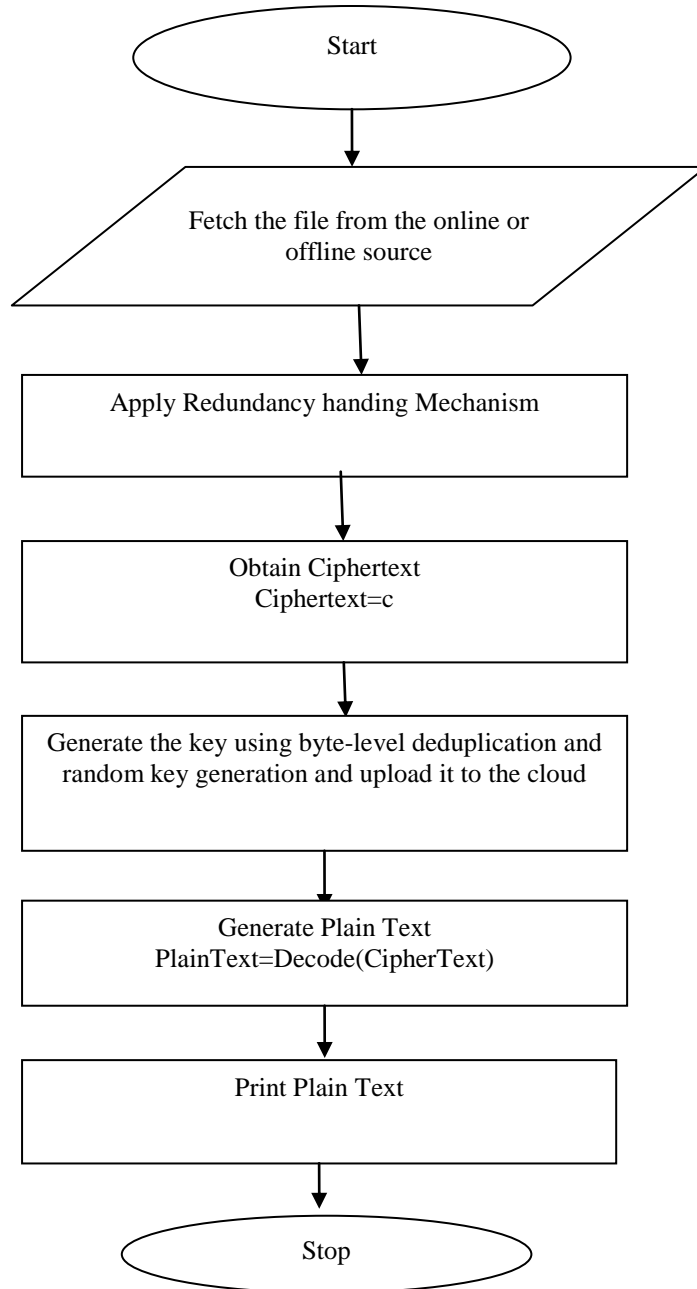
```
                    ┌──────────────┐
                    │    Start     │
                    └──────┬───────┘
                           │
                           ▼
              ╱─────────────────────────╲
             ╱ Fetch the file from the    ╲
            ╱  online or offline source     ╲
            ╲                               ╱
             ╲─────────────────────────────╱
                           │
                           ▼
              ┌───────────────────────────┐
              │ Apply Redundancy handing   │
              │ Mechanism                  │
              └─────────────┬──────────────┘
                           │
                           ▼
              ┌───────────────────────────┐
              │ Obtain Ciphertext          │
              │ Ciphertext=c               │
              └─────────────┬──────────────┘
                           │
                           ▼
              ┌───────────────────────────┐
              │ Generate the key using     │
              │ byte-level deduplication   │
              │ and random key generation  │
              │ and upload it to the cloud │
              └─────────────┬──────────────┘
                           │
                           ▼
              ┌───────────────────────────┐
              │ Generate Plain Text        │
              │ PlainText=Decode(CipherText)│
              └─────────────┬──────────────┘
                           │
                           ▼
              ┌───────────────────────────┐
              │ Print Plain Text           │
              └─────────────┬──────────────┘
                           │
                           ▼
                    ┌──────────────┐
                    │    Stop      │
                    └──────────────┘
```

Fig. 1

1.3 Generating Keys

In order to generate keys, the user must log in to the KSP. The files uploaded, are encrypted and corresponding keys are generated. The redundant files are neglected and rest of the files are uploaded with the public and private keys generated.

1.4 Encryption and Decryption

For encryption and decryption using byte level de-duplication algorithms are hybridized. The algorithm yield ciphertext after receiving files as plaintext. Verification of the overall procedure is in terms of time consumed and the size of the file that can be uploaded.

IV. RESULT AND PERFORMANCE ANALYSIS

The result is presented in terms of file size that can be uploaded. Reliability of encryption and decryption in terms of time consumed is also a performance metric. The comparison in terms of quality is given as under Space conservation is achieved using the proposed system. This space conservation is indicated through the following table

Table 1: Simulation results with different file sizes

Simulation	Block Level(KB)	Byte Level(MB)
Sim 1	1046	678
Sim 2	1055	500
Sim 3	2010	878
Sim 4	1034	645
Sim 5	1074	788

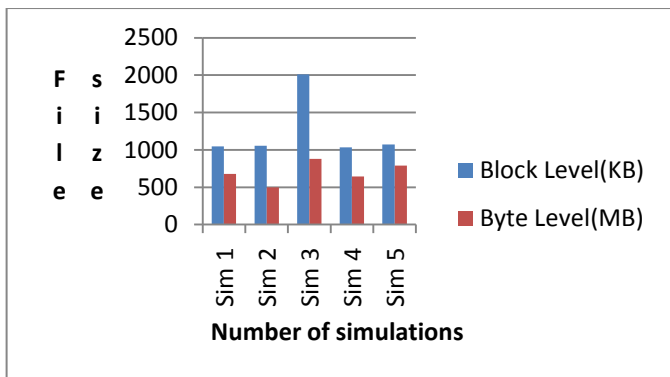The plots correspond to the existing and proposed literature is given as under

Figure 2: Existing and proposed file size comparison with the bar chart

The result deviation obtained using existing and proposed system is also given through the line chart. The line chart is generated by feeding the values obtained from the simulation as under
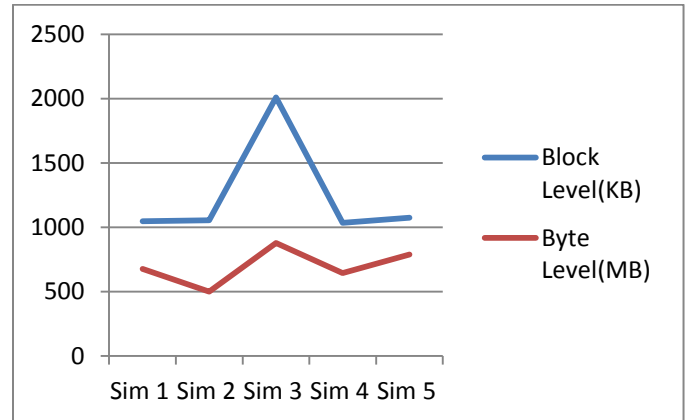

Figure 3: Line chart corresponds to space used by an existing and proposed system

The execution time subsequently reduced as file size to be uploaded reduced. The execution time is estimated by subtracting the end time of simulation from the start time of simulation. The comparison table for the same is given as under

Table 2: Simulation result of execution time

Simulation	Block Level(ms)	Byte Level(ms)
Sim 1	123	83
Sim 2	103	67
Sim 3	112	78
Sim 4	187	102
Sim 5	118	81

The plot for the execution time is given using bar and line charts. The chart generation is done using the facility provided through excel and jfreecharts.
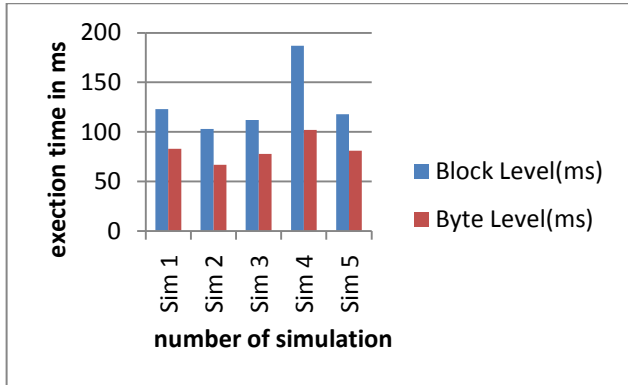
Figure 4: Execution time of simulation corresponding to the various simulation

The simulation results obtained are also plotted using the line chart. The proposed system execution time is better as compared to the existing system. The proposed system uses the byte level deduplication and the existing system uses block-level deduplication.
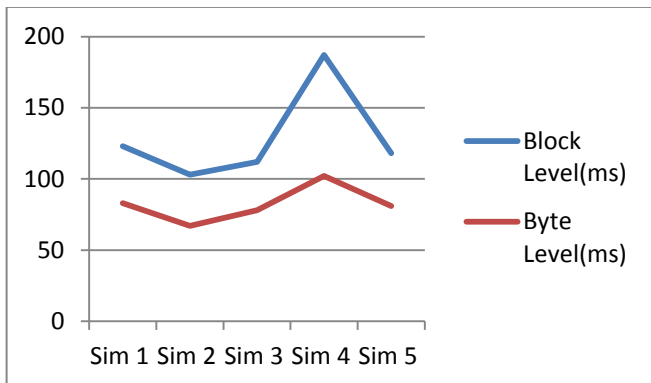

Figure 13: Execution time corresponding to the existing and proposed system.

V. CONCLUSION

Cloud computing not only provides the resources to the users but also give a big challenge of security. There are securities requirements for both users and cloud providers but sometimes it may conflict in some way. Security of Cloud depends upon trusted computing and cryptography. In our review paper some issues related to data location, security, storage, availability, and integrity. Establishing trust in Cloud security is the biggest requirement. The problems and corresponding solutions may require further investigation in terms of key size and complexity. The complexity of the key can be further enhanced by the use of a pseudo-random number generator within the key generation phase.
By incorporating complex key structure, cloud performance and user interaction can be further enhanced using complex keys and by reducing attacks.

REFERENCES

[1]   F. Sabahi, "Cloud Computing Security Threats and Responses," pp.

245–249, 2011.
[2]   X. Wu, R. Jiang, and B. Bhargava, "On the Security of Data Access Control for Multiauthority Cloud Storage Systems," pp. 1–14, 2015.
[3]   J. Aikat *et al.*, "Rethinking Security in the Era of Cloud Computing," no. June 2017.
[4]   K. Hwang, X. Bai, Y. Shi, M. Li, W.-G. Chen, and Y. Wu, "Cloud Performance Modeling with Benchmark Evaluation of Elastic Scaling Strategies," *IEEE Trans. Parallel Distrib. Syst.*, vol. 27, no. 1, pp. 130–143, Jan. 2016.
[5]   T. H. Noor, Q. Z. Sheng, L. Yao, S. Dustdar, and A. H. H. Ngu, "CloudArmor: Supporting Reputation-Based Trust Management for Cloud Services," *IEEE Trans. Parallel Distrib. Syst.*, vol. 27, no. 2, pp. 367–380, Feb. 2016.
[6]   M. Armbrust *et al.*, "A view of cloud computing," *Commun. ACM*, vol. 53, no. 4, p. 50, 2010.
[7]   R. Buyya, C. S. Yeo, and S. Venugopal, "Market-oriented cloud computing: Vision, hype, and reality for delivering IT services as computing utilities," *Proc. - 10th IEEE Int. Conf. High Perform. Comput. Commun. HPCC 2008*, pp. 5–13, 2008.
[8]   S. J. Nirmala, N. Tajunnisha, and S. M. S. Bhanu, "Service provisioning of flexible advance reservation leases in IaaS clouds," vol. 3, no. 3, pp. 154–162, 2016.
[9]   M. Marwan, A. Kartik, and H. Ouahmane, "Secure Cloud-Based Medical Image Storage using Secret Share Scheme," 2016.
[10]   D. V. Dimitrov, "Medical internet of things and big data in healthcare," *Healthc. Inform. Res.*, vol. 22, no. 3, pp. 156–163, 2016.
[11]   J. Li, J. Li, X. Chen, C. Jia, W. Lou, and S. Member, "Identity-based Encryption with Outsourced Revocation in Cloud Computing," pp. 1–12, 2013.
[12]   S. Seo, M. Nabeel, and X. Ding, "An Ef fi client Certi fi cateless Encryption for Secure Data Sharing in Public Clouds," pp. 1–14, 2013.
[13]   S. Wang, J. Zhou, J. K. Liu, J. Yu, and J. Chen, "An Efficient File Hierarchy Attribute-Based Encryption Scheme in Cloud Computing," vol. 6013, no. c, pp. 1–13, 2016.
[14]   D. Xu, C. A. I. Fu, G. Li, and D. Zou, "Virtualization of the Encryption Card for Trust Access in Cloud Computing," vol. 5, 2017.
[15]   A. Alabdulatif, H. Kumarage, I. Khalil, M. Atiquzzaman, and X. Yi, "Privacy-preserving cloud-based billing with lightweight homomorphic encryption for sensor-enabled smart grid infrastructure," *IET Wirel. Sens. Syst.*, vol. 7, no. 6, pp. 182–190, 2017.
[16]   J. Li, X. Lin, Y. Zhang, and J. Han, "KSF-OABE Outsourced Attribute-Based Encryption with Keyword Search Function for Cloud Storage," vol. 1374, no. c, pp. 1–12, 2016.
[17]   L. Jiang, D. Guo, and S. Member, "Dynamic Encrypted Data Sharing Scheme Based on Conditional Proxy Broadcast Re-Encryption for Cloud Storage," vol. 5, 2017.
[18]   C. Liu, S. Member, L. Zhu, J. Chen, and S. Member, "Graph Encryption for Top-K Nearest Keyword Search Queries on Cloud," vol. 3782, no. c, pp. 1–11, 2017.
[19]   C. Song, Y. Park, J. Gao, S. K. Nanduri, and W. Zegers, "Favored Encryption Techniques for Cloud Storage," pp. 267–274, 2015.
[20]   N. Veeraragavan, "Enhanced Encryption Algorithm (EEA) for Protecting Users ' Credentials in Public Cloud."
[21]   P. Xu, S. He, W. Wang, W. Susilo, and H. Jin, "Lightweight Searchable Public-key Encryption for Cloud-assisted Wireless Sensor Networks," *IEEE Trans. Ind. Informatics*, vol. XX, no. XX, pp. 1–12, 2017.
[22]   K. L. Tsai *et al.*, "Cloud encryption using distributed environmental keys," *Proc. - 2016 10th Int. Conf. Innov. Mob. Internet Serv. Ubiquitous Comput. IMIS 2016*, pp. 476–481, 2016.
[23]   A. El-yahyaoui, "A verifiable fully homomorphic encryption scheme to secure big data in cloud computing," 2017.
[24]   G. Thomas, "Cloud computing security using encryption technique," pp.   1–7.