ABE Based Secure Distributed Storage Scheme for Big Data in Cloud

Thesis submitted in partial fulfillment of the requirements for the award of degree of

Master of Engineering

in

Information Security

Submitted By:

Bhawna Jain

(Roll No. 801533003)

Under the supervision of:

Ms. Tarunpreet Bhatia

Lecturer, CSED

Dr. Anil Kumar Verma

Professor, CSED

COMPUTER SCIENCE AND ENGINEERING DEPARTMENT

THAPAR UNIVERSITY

PATIALA – 147004

September 2017

CERTIFICATE

I hereby certify that the work which is being presented in the thesis entitled, "*ABE based Secure Distributed Storage Scheme for Big Data in Cloud*", in partial fulfillment of the requirements for the award of degree of Master of Engineering in *Information Security* submitted in Computer Science and Engineering Department of Thapar University, Patiala, is an authentic record of my own work carried out under the supervision of *Ms. Tarunpreet Bhatia and Dr. Anil Kumar Verma* and refers other researcher's work which are duly listed in the reference section.

The matter presented in the thesis has not been submitted for award of any other degree of this or any other University.

(Bhawna Jain)

This is to certify that the above statement made by the candidate is correct and true to the best of my knowledge.

(Ms. Tarunpreet Bhatia)
Lecturer, CSED
Thapar University
Patiala

(Dr. Anil Kumar Verma)
Associate Professor, CSED
Thapar University
Patiala

ACKNOWLEDGEMENT

I have been waiting long for this moment to acknowledge all those who helped, motivated and believed in me during my thesis. I would like to thank my mentors **Ms. Tarunpreet Bhatia and Dr. Anil Kumar Verma,** who gave me guidance at each and every step, listened to my doubts and gave their valuable time for my success. I thank you for your extremely helpful suggestions and motivation at each step.

I express my profound gratitude to our director **Prof. Prakash Gopalan** for his inspiration and providing research environment at our campus. I extend my thanks to **Dr. Maninder Singh,** Head, Computer Science and Engineering Department, Thapar University for constant guidance during my research work and imparting self confidence in me.

I also want to thanks to my research committee members, non-teaching staff of the institute for their help and support and my fellow M.E. scholar Ms. Shruti Sachdeva for providing help during my thesis documentation.

At last, I also want to thanks to my family and friends for their emotional support.

ABSTRACT

Cloud computing is an emerging field of research. Data security and privacy both are crucial issues that may become an obstacle in various cloud applications. Cloud computing consists of retrieving data and storing programs over the internet rather than processor's hard drive. The cloud is just a symbol for the Internet. There are various attractive offers given by the cloud vendors for storage service. They provide scalable and gigantic storage space for its users like Drop box, Amazon, Google Drive, and Microsoft's One Drive. Security is one the major issue in the cloud. Cloud service providers want to secure their infrastructure and software platforms from the hackers. The users ensure that provider has taken proper security action to protect their data. So there is a need to enhance security of the cloud. There are various encryption techniques available like RSA, Blowwfish, XOR, DES, AES etc. Every technique has its own advantages and disadvantages. The existing techniques are not that much good for security purposes. To overcome this problem, a novel technique has been proposed based on three different keys. In proposed work, different keys have been used for the encryption and decryption purpose. In the proposed work, ABE based secure distributed storage scheme for big data in the cloud (ASDSS) has been proposed to provide security for the massive distributed data for big data in different clouds. In ASDSS technique, KP-ABE scheme has been used for encryption and decryption purposes. CloudSim is used for simulation purpose. Experimental results shows that proposed scheme is secure and takes less time for data storage as well as data retrieval. For the performance analysis, we have compared ASDSS technique with SA-EDS, XOR, RSA and Blowfish. Proposed technique is more efficient in terms of data storage and data retrieval as it takes less time than other techniques in both parameters.

Keywords: Big Data, Security, Encryption, Decryption, SA-EDS, XOR, ASDSS, RSA

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

ABE	Attribute Based Encryption
AES	Asymmetric Encryption Standard
ASDSS	ABE based Secure Distributed Storage Scheme
MDS	Massive Data Storage
NDP	Named Data Packets
PNL	Pre-Stored Named List
XOR	Exclusive-OR

CHAPTER 1

INTRODUCTION

1.1 Introduction to Cloud Computing

Cloud computing consists of data retrieval and program storage over the internet rather than computer's hard drive. For the internet, cloud is just a symbol. To accumulate data and run programs from the drive is known as local storage and computing. It simply defines that retrieving your data is easy and fast from one computer to other on networking as shown in Figure 1.1.

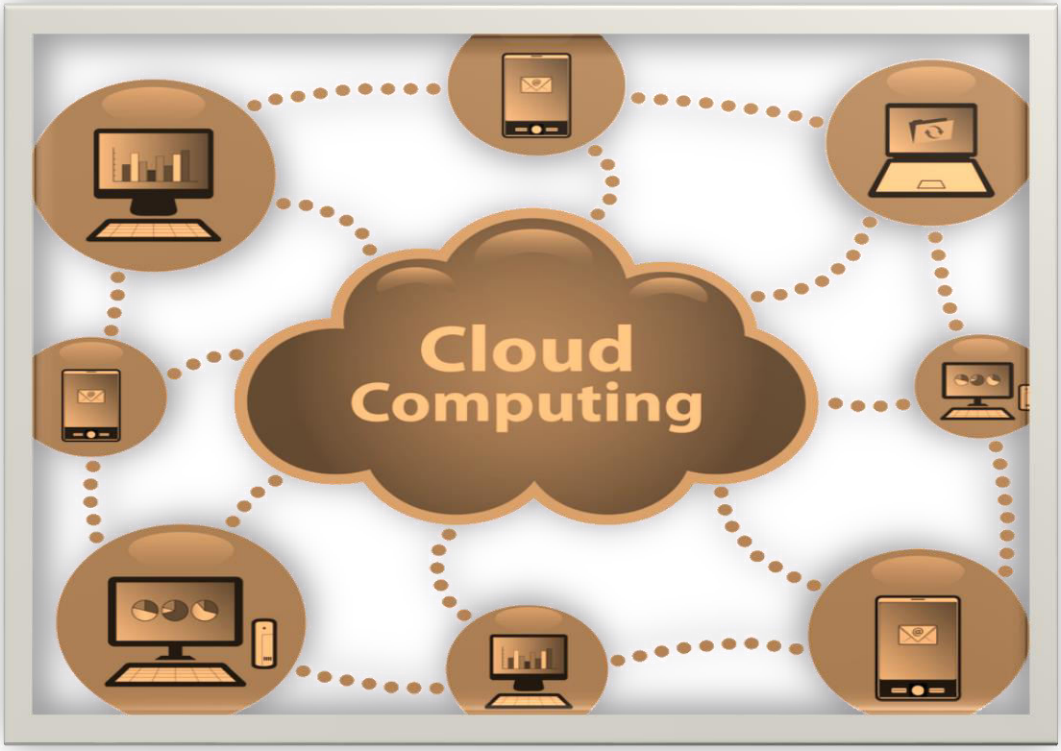

Figure 1.1: Overview of Cloud Computing

The main services of hybrid technologies like Dropbox, Box and Sugar Sync works for cloud because they collect an online synced type of your files, Further, they synced these files with local data storage. Synchronization is a keystone of the cloud computing experience.

Moreover, it is also concluded that cloud computing, is a community of people having dissimilar devices and they all need same synced data.

A decade ago, IT mission or start-up that wanted dependable and internet linked up with assets for computing for several facts centers. Today, it is straightforward to lease computing time and garage of any size. The variety begins with sensible machines slightly powerful enough to serve internet pages to the equal of a small supercomputer. At the identical time, cloud services and sources are distributed over the world. This setup guarantees a high obtainability and unimaginable durability.

1.2 Cloud Architecture

With the passage of time, three main cloud architecture models have been recognized. They all share the same resources and to that conclusion usually, virtualize computing and abstract storage layers.

a) Private Cloud

Velocity, variety, and volume are exponentially growing, so, the infrastructure of the enterprise must have to adapt it, also it will become progressively agile, efficient and scalable if organizations will remain competitive. To complete their needs, Synnex Corporation can consider a lead distributor of IT solutions and products which have announced that; this is an initial distributor to offering fully combined, big data turnkey private cloud appliance powered by Nebula to channel. These appliances are private cloud system fully collective engineered to deliver workloads for both MongoDB and Apache Cassandra, enables open and elastic infrastructure to save and manage big data. The latest rack system including industry-standard servers, MongoDB, Apache Cassandra, Nebula One Cloud Controller, backed by SYNNEX' powerful distribution model. These big data appliances allow an open and elastic infrastructure for storing and managing big data. As open source system NoSQL database technology, Apache Cassandra is providing big data multi-site distributed computing across various data centers, during MongoDB is cross open source database system document-oriented [24].

b) Public Cloud

Public cloud is sharing resources for storage, data transfer, and processing. Hence, the user has visualized private computing, isolated storage, and environments. Security concern that entices few to adapt custom deployments or private clouds is

the broad amount of projects and customers irrelevant. Visualization provides access to some other user' which is extremely difficult for data. In the real world, many problems are around public cloud computing which can be more mundane such as fluctuating performance and data lock-in of individual instances. The data lock-in is soft measure and worked with making data inflow to cloud provider very cheap or free. The copied data out of other local systems or other providers can often be more expensive. It is a practice that encourages utilization and not an insurmountable problem, more services from cloud service provider instead to move data out and in for different processes or services. Usually, it is not sensible with complexities and network speed around dealing with various platforms [25].

c) Hybrid Cloud

The size of organization or industry doesn't matter, chances are business find value in big data. But, how you can manage this data? Is there an optimal environment to process and store huge amount of data in the cloud? Hybrid cloud can hold their answer. Their dedicated public and private cloud combination offer optimal performance, enhanced security and better financial savings for business working with big data. Moreover, business's big data can include customer's financial data, contact information or social network history. Regardless of the type of data, security can paramount. Most of the hybrid clouds are offering protection to provide security services that can be found in the traditional dedicated environment like Intrusion Prevention System (IPS), Web Application Firewalls (WAF), Security Information and Event Management (SIEM), File Integrity Monitoring (FIM). It creates an environment which allows adding layers of security which can be more difficult to found at providers that provides public cloud solution only[26].

1.3 Introduction to Big Data

The growth of cloud data stores and cloud computing has been expediter and predecessor to the appearance of big data. A Cloud computing is co-modification of data storage and calculating time by means of identical technologies. It has important benefits over conventional physical deployments. Nevertheless, cloud platforms

originated in numerous forms and from time to time it is combined with traditional architectures.

It carries the impasse for decision makers in the responsibility of huge data projects. In what way and which cloud computing might grow the best choice for computing requirements, particularly if it is the project of big data? These preparations often show bursting, variable or storage requirements and enormous computing power as shown in Figure 1.2, in which big data and its features are displayed. It can store a large amount of data. At the same time, profitable stakeholders visualize cheap, swift and independent project outcomes and products. This article is presenting cloud storage and cloud computing, the essential cloud architectures, and deliberates what to aspect for and how to get happening with cloud computing.

Figure 1.2: Big Data

1.4 Big Data Challenges

Big data is resolving numerous existing problems concerning the high amount of data, it is a continually altering area that is constantly in growth and still poses some matters. In this section, we are signifying some of the issues not yet talked by cloud computing and big data. As the quantity of data develops at a quick rate, keeping all the data is bodily cost-ineffective. Although, corporation's needs must have to make a rule to label their cycles of life and expiry date of data. Furthermore, they must describe who is retrieving and with what purpose clients' data can be reclaimed. Such

as data transfer to the cloud, privacy and safety develop a concern which is with research. Big data DBMSs typically deal with lots of data from various sources, and as such heterogeneity is also a problem that is currently under study. Other issues being investigated currently are load balancing, resource allocation, scalability and migration to cloud etc.

- **Privacy:** Data Harvesting and use of logical tools to excavation information raise various privacy worries. To confirm shield privacy and data safety has become enormously problematic as data is pretended and range around the sphere. Analytics frequently mine users' delicate data such as their medical proceedings, energy depletion, supermarket records, online activity etc. This data is visible to inspection, hovering anxieties about perception, summarizing, elimination and loss of control [20]. Traditionally, corporations used multiple approaches of de-identification to distance data after actual identities. Although, it is demonstrated that when records were anonymized, it could nevertheless re-diagnosed and credited to specific individuals. A way to remedy this trouble was to deal with all facts and situation to a regulatory framework. Privacy and data protection laws are premised on man or woman manipulate over records and on standards such as information and motive minimization and trouble. Nevertheless, it isn't always clear that minimizing records series is usually a realistic technique to privacy. Nowadays, the privateness processes when processing activities seems to be primarily based on person consent and at the information that people intentionally provide.

  Privacy is certainly a difficulty that needs similarly improvement as systems work big quantities of personal facts every day.

- **Heterogeneity:** Big information concerns large volumes of facts but special velocities and wonderful range. The latter comprehends massive and heterogeneous quantity of data comes from various independent assets. Variety is one of the "most important components of huge records characterization" [21] that is prompted with the aid of perception which stores all forms of records can be beneficial to both a commercial enterprise and technological know-how. Data comes to large facts DBMS at exclusive velocities and codes from a couple of assets. This is due to the fact that one of

kind information creditors decides on their very own schemata or protocols for statistics file, and nature of different packages results in diverse facts. Dealing with such extensive form of records and special velocity cost is a hard task that Big Data structures should take care of. This venture is annoyed by means of the truth that new kinds of the documents are continuously being created with none form of standardization. Though, supplying a constant and widespread manner to represent and explore complicated and evolving relationships from these records nevertheless poses a venture.

- **Data Governance:** Space of data storage is inexpensive and it can be reduced further with the help of hardware devices. Conversely, big data DBMS also focus on other expenses such as infrastructure, energy, maintenance, and software licenses. By combining all these expenses, total cost can be estimated which is seven times higher than hardware acquisition cost.
  In the equal amount of proportion big data grows. This growth is under control.

  To eliminate various problems that occurs during big data policy, Data Governance came into limelight. Various real time and organizational policies are discussed in detail that how economic life cycle data is managed. These practices comprise of three different categories:

  1. Key IT and non-IT decision makers and their roles and responsibilities are identified by the Structural practices and provide information regarding data ownership, cost management, and value analysis.
  2. Data Governance policies are applied according to the Operational practices. Normally, data retention, data migration, cost allocation, access rights, and backup and recovery these policies span a variety of actions [22].
  3. It also defines links of the CIO, data users, business manager for relational practices formally describe in terms of value analysis, knowledge sharing strategic IT planning, and education and training.

Data Governance is a popular time period which applies to businesses with huge data sets that define rules to keep treasured information as well as to control the facts accesses at some stage in their life cycle. It is a matter to

address cautiously. If policies of the government are not enforced, it is clear that they are not followed by us.

- **Disaster Recovery:** Data is a lot precious commercial enterprise and dropping statistics could be in reality resulting in losing price. In case of dangerous or emergency accidents which include floods, earthquakes, and fires, statistics losses want to be minimal. To fulfill this requirement, in case of any incident, information should be quick with minimal downtime and loss. However, despite the fact, that this is a completely critical difficulty, the research in this precise place is distinctly low [23]. For large corporations, it's miles authoritative to describe a disaster restoration plan – as part of the information governance plan – that no longer simplest is predicated on backups to reset records but also in a fixed of strategies that permit short replacement of the misplaced servers [23]. From a technical angle, the work described in [23] gives an awesome methodology, offering a “multi-reason method, which permits data to be restored to multiple websites with multiple techniques", ensuring a recuperation percent of just about 100%. The examiner also states that normally, information recuperation strategies use what they name a "single-basket approach", because of this there is handiest one destination from which to comfortable the restored facts. As the lack of statistics will potentially bring about the loss of money, it's far essential so that you can respond effectively to dangerous incidents. Successfully deploying massive facts DBMSs inside the cloud and retaining it usually to be had and fault-tolerant may strongly rely upon catastrophe recuperation mechanisms.
- **Data Management and Resource Allocation:** Resource allocation is the current and advanced idea of cloud computing. It is the most challenging issue in data center network. During revenue exploitation, virtual resource is optimized. Furthermore, power consumption is the issue for data centers.
- **Load Balancing:** It is related to download performance and storage utilization. This problem mainly occurs in distributed nodes.
- **Availability and Scalability:** Performance degradation and oversizing problems occur due to unpredictable in the cloud.
- **Compatibility and Migration to Clouds:** The significant concerns during the procedure of migration such as increasing dependency of enterprises on

outside third party, departmental downsizing, lack of understanding about cloud features and structure, lack of supporting resources and uncertainty in new technology.

- **Interoperability and Communication between Clouds:** At each level of communication and interoperability have some shortcomings that make the procedure of attaining interoperable cloud computing environments more challengeable.

Despite these huge benefits, there are some worries and experimentations related to the new technology. The most significant matter is connected to security and privacy issues in cloud-based environments.

1.5 Motivation of Thesis

Cloud computing data security is the subcategory of information and cyber security. Data is stored on the variety of servers. Clients must trust the cloud servers on data availability and data security as well. To maintain the reputation, many times service providers have to hide the data integrity from the clients. To avoid hiding problem, third party auditor is introduced to ensure data owner that their data is safe on the cloud. Cloud users save their data in cloud servers so that data reliability, as well as data security, is the primary concern. There are various algorithms to enhance the security of big data in the cloud. These algorithms are applied according to the requirements and security level, so that data cannot be hacked easily. So, to make the network secure, security should be provided.

1.6 Overview of Thesis

The main objectives of this research are to provide security to the big data in the cloud. In Chapter 1st, Introduction to Cloud Computing, Cloud Architecture, and Big Data have been discussed. We have also discussed big data challenges in this chapter. In Chapter 2nd, the Literature Review related to the Cloud Computing, Big data, and their security, features, defects, and enhancements have been discussed. In Chapter 3rd, Big data in the cloud, its security, various existing Encryption algorithms have been discussed. In Chapter 4th, Problem statement, research gaps, and objectives of the study has been discussed. In Chapter 5th**,** the Proposed scheme has been discussed.

We have described ASDSS with details of computational software used for the work. In chapter 6th, Results of proposed scheme has been discussed. In chapter 7th, conclusion and future scope of the proposed scheme have been discussed.

CHAPTER 2

LITERATURE REVIEW

Review of Literature is an evaluative report of studies found in the literature related to our selected area and our topic is the security of big data in the cloud. The review describes, summarizes, evaluates and clarifies this literature. It is giving a theoretical basis for the research and helps us to determine the nature of our own research. We do a literature review to ensure the thorough understanding of the security of big data in the cloud, to identify potential areas for research and the similar work done within that area.

Li, Yibin et al. [23] describes how many cloud applications has been restricted because of the critical issues of data security and privacy and one of them is sensitive data is reachable to the cloud operators. So an intelligent cryptography approach is proposed in this paper with which the partial data is not directly reachable to the cloud service operators. In this proposed scheme, the distributed cloud servers are used to store the data after dividing the data file and the approach used is SA-EDS model, is best for the proposed method which is comprising of AD2 Algorithm, Secure Efficient Data Distributions (SED2) Algorithm and Efficient Data Conflation (EDCon) Algorithm.

Zhao et al. [24] a novel security design are proposed in for G-Hadoop depends on the security solutions such as SSL protocol and public key cryptography designed dedicatedly for distributed environments. The job submission process and users authentications are simplified by this security framework. The traditional attacks protection is provided to the G-Hadoop system by different security mechanisms in the designed security framework.

Manogaran et al. [25] proposed a Big Data protection by MetaCloudDataStorage Architecture in Cloud Computing Environment. The user's number that logged into the cloud data center is found by using Map Reduce framework and the various data elements map to providers is protected by using the proposed framework that is a MetaCloudDataSortage interface. High implementation effort is required by this

proposed scheme providing cloud computing environment valuable information.

Liu et al. [26] delivered an investigation on cloud authenticator-based data integrity verification techniques and things data internet. Simple aspect in research problem is analyzed. First, the research motivations and methodologies are summarized for illustrating the research problem. Secondly, the representative approaches several current achievements are summarized and compared. Finally, future developments view possibilities are introduced.

Ramachandran et al. [27] introduced a general evaluation of the cloud and data security literature. The data security and security design are provided as a service which is explained in the use of Business Process Modelling Notations (BPMN) and the two cloud service providers examples with their security design are presented in this paper. The attack on the security service section can be identified by using the BPMN in case of any security breach. For more resilient and reliable security services in business, integrating CCAF version 2 with BPMN is needed.

Baek et al. [28] proposed a secure cloud computing based framework called as Smart Frame. Cloud computing centers hierarchical structure is built in their framework which provides cloud services of a different type for information management and big data analysis. By addressing critical security issues in the proposed framework, a security solution is provided which is based on identity-based encryption, signature and proxy re-encryption.

Sookhak et al. [29] proposed an efficient Remote Data Auditing (RDA) technique for cloud storage system which is based on algebraic signature properties that incur least computational and communication costs. Divide and Conquer table (DCT), a new data structure is also presented which can provision dynamic data structure efficiently such as insert, append, modify and delete. In comparison between another state of art RDA techniques and their proposed approach shows their approach is highly efficient and secure which helps in reducing the communication and computational charges on the server and auditor.

Gai et al. [30] focused on big data issues and its practical implementation is

considered in cloud computing. To maximize the privacy protections efficiently, the approach designed is Dynamic Data Encryption Strategy (D2ES). DED algorithm is what mainly support D2ES model that was developed for encryptions of the dynamically alternative data package having different time constraints. The main aim of this approach is maximizing the privacy protection with the usage of choosy encryption strategy with the specific execution time requirements.

Pasupuleti et al. [31] proposed a well-organized and protected preserving privacy method which is used for resource- constrained mobile devices data outsourcing in the cloud computing and for encryption of the data, a probabilistic public key encryption algorithm is employed. For file retrieval from the cloud over the data encryption, the ranked keyword search is invoked. The aim of this approach is to achieve a data encryption system that is efficient without data privacy sacrifice.

Sood et al. [32] proposed specialized procedures and different techniques in a frame work in which from beginning to end the data is being protected efficiently, i.e., from the owner to the cloud and after that to the user. The three cryptographic parameters that the user presented based on commence of the data classification i.e., Confidentiality (C), Availability (A) and Integrity (I). Also follows the plan for the information protection which utilizes various measures such as the SSL (Secure Socket Layer) 128-bit encryption and which can be elevated up to 256-bit encryption as per required, for integrity data check using MAC (Message Authentication Code), searchable encryption and for cloud data division into three sections in cloud. The supplementary protection is rendered by the data divided into three sections and data simple access.

Shaikh et al. [33] described that how in cloud computing, the active area of experimentations and research is data security and privacy. The organizations which are moving onto the cloud, protection of the data privacy and data leakage is crucial. There are various types of data and protection degree required for every different type of data is variable. A classification technique is proposed in which on various dimensions bases, parameters are defined. On the level bases, data security is provided in which protection is required and on data set classification bases at the storage security provisions are applied.

Li et al. [34] focused on the problem of the security issue and proposed a technique for secure commercial services in cloud computing on multimedia big data that novel approach is Semantic-Based Control (SBAC). The entitled proposed approach is *IntercroSsed Secure Big Multimedia Model* (2SBM). This approach is basically designed for secure accesses of the various media through multiple platforms provided by the cloud. The proposed model is supported by the main algorithms which include *Ontology-Based Access Recognition (OBAR) Algorithm* and the *Semantic Information Matching (SIM) Algorithm.*

Qiu et al. [35] proposed an algorithm that is called Proactive Dynamic Secure Data Scheme with Attributed-Based Access Control, as well as data self-deterministic scheme in, a financial customers' private information and its aim, is that the private data do not get unanticipated by the third party. The proposed scheme is supported by two main algorithms that are Attribute-based Semantic Access Control (A-SAC) Algorithm and Proactive Determinative Access Algorithm. This papers main contribution is having three aspects that are: first, for constraining data accesses a semantic approach is proposed. Second, a user-centric approach is proposed that prevent the users' data proactively from processes that are unexpected on the cloud side and finally, a higher-level secure sustainability is in the proposed scheme as it deals with dynamic threats including the future hazards and the emergence.

Chandrasekaran et al. [36] described that the IBE scheme where identities are used as a characters string. On three theories, its working principle depends on bilinear pairing, quadratic residue, and lattices. The extended type of Individuality based encryption in which expressive attributes set is utilized instead of identity. It is an access control mechanism which is efficient for computing cipher text for the user group that is based on access structure. Most often on bilinear pairing, the ABE existing schemes are constructed. The constructions of a novel ABE scheme takes place on quadratic filtrate and the quality union on the fundamental arithmetic theorem bases. In this approach, in the cloud environment, the big data gravest threat is unauthorized user access control which is prevented owing to fewer occurrences of the set of a user attribute in a squared value.

Kang et al. [37] proposed a novel approach so as to place the data in the in cloud

storage systems. First, a linear programming model as a data placement problem is formulated so as the data's total retrieval time behavior and over storage nodes, it is divided and distributed under the security constraint. The problem is solved by developing a heuristic algorithm for cloud storage Systems (SEDuLOUS) that is Security-aware Data placement mechanism. And the proposed algorithm effectiveness is demonstrated through comprehensive simulations. Significantly the proposed algorithm shows the simulation results reducing the recovery time up to 20% for the random-network-topology systems and for the Internet2-topology system it is 19% as compared to baseline approaches with only the security requirement consideration.

Gai et al. [38] explained the decision tree techniques are being utilized for the prediction of sharing of data potential risks between financial service institutions. The approach intended is to decrease the confidentiality leakage chance which shares the multiple datasets. The Supervised lEarning-Based Secure Information Classification is a proposed model which is supported by a proposed algorithm that is DTRP *algorithm*. Performance of the procedure used is good as shown in experimental evaluations in accuracy examinations but causing the additional computation workload. In this paper, the issues it focuses on and the approach it proposed uses the combination of supervised learning techniques for the classification of the information so as to avoid releasing of harmful information either financial service providers or customers.

Gai et al. [39] described that there is a remarkable increase in the difficulties of justifying losses for financial firms from cyber incidents and that has derived the rapid growth in the *Cybersecurity Insurance* (CI). In the current applications, there are some uncovered number of dimensions as still CI is at its stage of exploration and one of the critical issues in the CI is cyber-attack. In this paper, the focus of the CI operations is on the offered cloud based facility and a safe cyber event analytics framework is proposed by means of big data. This new framework is CA-HCIA in which the CI cost is reduced without the security level lowering down.

Chen et al. [40] explained the data security and privacy security issues, the sensitive data split-up and access control fundamental challenges. The designing of the unified identity management and privacy protection frameworks set is the main objective of

applications or cloud computing services. The privacy protection mechanisms bases accountability will be achieved dynamically and informs real-time, approval and checking for the data possessors as the private data is retrieved. The data safety and privacy protection matters concise are provided in this paper associated with cloud computing and also discussed some current solutions.

Baker et al [41] described that in cloud network environment, the high priority aim is the energy efficiency which includes sustainable data centers. In this paper rather than with data centers energy consumption, it dealt with cloud routing energy consumption. And evaluated and proposes dubbed GreeDi which is a new energy efficient routing framework and situation calculus gives a cloud network connectivity formal analysis. The physical Italian ISP topology is what on which GreeDi algorithm is evaluated and three dissimilar routes are there for the data centers of the green cloud.The evaluation result gives the greatest energy effective route decision that can only be predicted after each successful transmission of data making sure that the traversed nodes based calculations are done.

Fazio et al. [42] aimed hybrid architecture in which Document and Object oriented strategies are coupled for optimizing the data loading, inquiring and recovery. In this paper, an architecture scheme is presented and implementation details are discussed in architecture development within a specific case use. The big data storage issues are dealt in paper owing to observe activities in smart environments. Then, the storage technologies of different -types are integrated into the new storage solution that is proposed.

Demirkan et al. [43] proposed a conceptual framework in the cloud for DSS after the requirements list for service oriented DSS is defined, and research directions are discussed. In this paper, a unique contribution is its perception that how the invention concerned with DSS environment is servitude, and the chances and experiments are demonstrated in the cloud on engineering service oriented DSS. A scale, scope and speed economy is enabled by DSS in CLOUD. In service science, new knowledge is contributed by this article in which for a broader audience information technology policy perspective is tied to the perspectives of database and design science.

Teli et al. [44] proposed an online algorithm for finding out optimal cost data aggregation site which is among the geographically distributed data centers. An optimal cost solution is given in the proposed approach from different geographically distributed data centers for the data aggregation that can process efficiently at a single site using distributed frameworks. A graph model is proposed that are of Geo-distributed data centers. Better results are obtained by the proposed approach as shown in the results obtained in the online cloud environment.

Vennila et al. [45] introduced a Parallel Symmetric Matrix-based Predictive Bayes Classifier (PSM-PBC) model for efficient computation of the Big Data and sharing information in Cloud environment and constructing a Tridiagonal Symmetric Matrix initially on Big Data that is distributed in parallel is explained. The data computation rate increase is enabled using a Householder transformation. The prediction rate is improved after the real-value diagonal search data is evaluated by Cross-Validated Bayes Classifier. Finally, the efficient predictive analytics is provided by the MapReduce function on Bayes Classes regarding Big Data.

Chang et al. [46] examined a comparison of setting process and the used metric is discussed. Comparative analysis of the performance between cloud and non-Cloud systems, its biomedical scientist deployment conducted for identification of efficiency and performance improvement has been discussed. Before, Organizational Sustainability Modeling (OSM) is used, for ensuring the achievement of the fair comparisons during and after the experiments. Control rates of actual and expected execution time are defined by the OSM and are used for understanding related key outputs to both Cloud and non- Cloud experiments. Under these two case studies, 40 experiments are done on both Cloud and non-Cloud systems. The focus of the first case study is on shifting and backing up 10,000 files of 1 GB each. And the emphasis of the 2nd case on shifting and backing up 1000 files 10 GB each.

Zhang et al. [47] introduced a hybrid approach and examined on the cloud rather than big data and proposed a scheme which combines together the Top–Down Specialization (TDS) and Bottom–Up Generalization (BUG). Automatically, the hybrid approach selects one among the two mechanisms by the judgment of the user-specified *k*-anonymity parameter and work load balancing point. In a highly scalable

way, both TDS and BUG are accomplished via Map Reduce jobs series that is designed deliberately. As demonstrated in the performance evaluation results that significantly the hybrid approach improves the sub-tree data anonymization scalability and efficiency in compared with approaches that are already existed.

Assuncao et al. [48] discussed Big Data applications analytics that are carried out on Cloud various approaches and environments. Four important area that it revolves around are of analytics and Big Data, namely (1) data administration and associate architectures; (2) model development and scoring; (3) visualization and user interaction; and (4) business models. In technology, possible gaps are identified through a detailed survey on Cloud-supported Big Data computing and evaluate its performance.

Malekabadi et al. [49] discussed the challenging issue of the data storage and processing. The main aim of this study is to present a model based on NoSQL database for healthcare data storage. The survey on the health data nature selects documents-based DBs despite the NoSQL database of various types. In this paper, comparison with the previous data model is evaluated in which inquiry time, preparation of data, flexibility, and extensibility parameters are considered. In the Cloud environment implementing the present model for the distributed properties access and Shard property is applied for data distribution.

Malhotra et al. [50] described the worlds hot topic nowadays is cloud computing in which information can be accessed by the customers via a web browser. Evaluating the cloud environment performance is critical to cloud computing adoption and evaluation increases. In this paper, the cloud computing simulators that already exist are reviewed up to their best knowledge. Furthermore, cloud computing simulators of two types that exist are indicated, that is, software based simulators and together with software and hardware based simulators. Finally, comparing and analyzing the features of the existing cloud computing simulators.

Xiong et al. [51] discussed the challenging task of obtaining the full lifespan privacy safety, access control after sensitive data sharing on cloud servers is not achievable. For tackling this problem, key-policy attribute-based encryption scheme with time-

specified attributes (KP-TSABE) is planned which is an original secure data self-destructing scheme in cloud computing. In the KP-TSABE scheme, a time interval is considered in every cipher text while private key and time prompt are connected. If key's access structure is pleased by the cipher text with its suggestion with features and in an allowable time interval, both times instant is then only decryption of the cipher text takes place. Resolve approximately significant security problems are solved by using KP-TSABE which repairs user distinct approval period and brings fine-grained access control.

Horvath et al. [52] explained a making the access control more appropriate for the data storage in the cloud using attribute-based encryption. They proposed that the given full resistor to the encryptor over the admittance rights so as key managing that is feasible is provided in multiple independent authority cases which are enabled the viable user revocation. The arrangement has official security proof given as in general bilinear group and random oracle models.

Chang et al. [53] explained the review on the CCAF reason and components for protecting the information safely and based on the necessities, the device layout of the CCAF is illustrated and the CCAF multi-layered security implementation is demonstrated. The BPMN simulation use which lets in the evaluation of the selected protection performances earlier than the actual implementation. Demonstrating on this paper the CCAF multi- layered protection safety for the facts protection in real time. The safety 3 layers it has: 1) firewall and get admission to manipulate; 2) identification management and intrusion prevention and three) convergent encryption.

Han et al. [54] planned a privacy-preserving decentralized key-policy ABE arrangement where secret keys are delivered by each authority to a user independently. So, if corrupted are the numerous authorities, the user's qualities are not composed by tracing GID. Particularly, standard complexity molds are compulsory in the scheme and among the numerous authorities, any collaboration is not required in contrast with the previous similar scheme in which non-standard complexity expectations are required and multiple authorities' interactions. Based on standard complication assumptions with privacy-preserving, it is the first decentralized ABE scheme.

Li et al. [55] proposed a novel patient-centric framework and for data access control as a devices suite to PHRs which is deposited in semi-trusted servers. For PHRs file, fine-grained and scalable data access control is attained and for encrypting each patient's PHR files leverage ABE techniques. This paper emphases on the numerous data owner set-up, and divides the PHR system user into several security domains which decrease the key management difficulty importantly for owners and users. Access policies or file attributes dynamic alteration is also allowed in the scheme in which well-organized on-demand user/attribute cancellation is supported and also under alternative break-glass access is supported.

CHAPTER 3

BIG DATA SECURITY IN CLOUD

3.1 Need of Security

Cloud computing is providing wide storage space, network access, resources, operator and applications [1]. It is shared network that delivers services over real time environment. Cloud computing has various amazing features like on- demand user, shared pool computations, pay as per use, on-demand services, rapid elasticity etc [2]. Cost saving is the main advantage of the cloud. Security of cloud is a major disadvantage of the cloud. Information stored on the cloud server is the biggest challenge of cloud [3]. The User is not comfortable to handle the data over the cloud. To make server secure, various encryption techniques have been developed using passwords [4]. Also, bi-directional security is needed in the cloud. Cloud service provider wants to secure their infrastructure, software, a platform from hackers and user ensure which provider has taken proper security action to protect their data. Therefore security is needed at the both sides [5]. To protect the data and server from the hacker is called as information security. To prevent loss of data and modification of data comes under information security. So, cryptographic techniques are used to protect the data from security hacks. Cloud computing data security is the subcategory of network and information security. Data is stored on the variety of servers. Clients must trust on the cloud servers on data availability and data security as well. To maintain the reputation, many times service providers have to hide the data integrity from the clients. To avoid hiding problem, third party auditor is introduced to ensure data owner that his data is safe on cloud [6]. Data reliability, as well as data security, is the primary concern, therefore, cloud users save data in cloud servers.

Cloud data storage with the untrusted server, Data integrity verification is one of the biggest concerns. Cloud computing is a suitable model for on-demand network access. As we discussed above, cloud computing has three service models.The storage is distributed in nature and has allowed the mass remote data via Storage-as-a-Service (StaaS) model and it is one of the significant technologies used in cloud Computing. With the advent of technology in web Services and networks, this approach is highly acceptable [8, 9]. There are various attractive offers given by the

cloud vendors for storage service. They provide scalable and gigantic storage space for its users like Drop box, Amazon, Google Drive, and Microsoft's One Drive. Security is one the major issue in the cloud.

3.2 Classification of Algorithms

Encryption algorithms for the security can be divided into 2 categories according to their behaviour. These categories are:

a) Symmetric
b) Asymmetric

In symmetric, for encryption and decryption the same key can be used but in asymmetric keys can be different for encryption and decryption purposes.

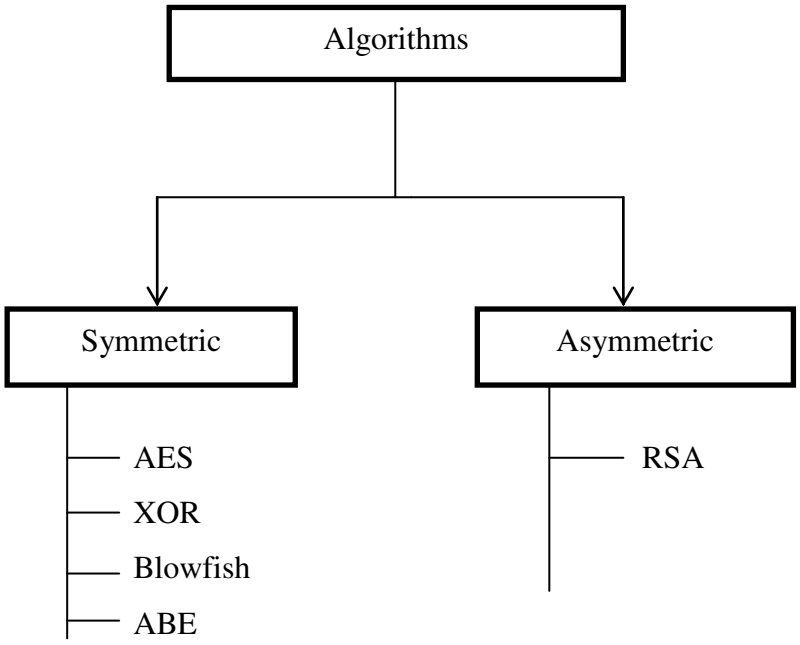

Figure 3.1: Classification of Algorithms

According to Figure 3.1, AES, XOE, Blowfish, and ABE comes under symmetric algorithms and RSA comes under Asymmetric algorithms.

3.3 Introduction to Some Existing Algorithms for Encryption

There are various algorithms present for encryption. These algorithms are:

a) Attribute Based Encryption (ABE)

Nowadays we can detect the spread of dispensed programs where sensitive information has to be shared with a couple of events. To be able to facilitate these distributed systems, a community infrastructure can be utilized to allow shared storage (e.g. In the cloud) and access to the programs' assets. Despite the fact that this new paradigm presents a number of advantages with the aid of disposing of organizational boundaries and increasing operational flexibility, it requires further security mechanisms that want to preserve sensitive data concerned. Not like in ordinary situations, it encrypts a message; allotted (collaborative) methods require extra flexibility. The access to their assets is regulated, permitting entry to events that satisfy access coverage alternatively than to a specific set of parties [16].

In an attribute-established encryption process, at least one cipher text is needed to encrypt data with cryptography keys. Alternatively, both customers' confidential keys and cipher texts might be associated with a suite of attributes or a policy over attributes.

A consumer is in a position to decrypt a cipher text if there is a “match” between their personal key and the cipher text. Sahai and Waters provided a threshold ABE procedure where cipher texts were considered with a set of attributes S and a consumer’s individual key was connected to each of the threshold parameter k and another set of attributes S. In order for a user to decrypt a cipher text at least k features must overlap between the cipher text and their individual key. One of the foremost common motivations for this was to plan an error-tolerant (or Fuzzy) identification-founded encryption scheme that might use biometric identities[17].

The main problem in this algorithm is that data owner has to use every key for encryption. Further, it is divided into two categories:

1. Key Policy ABE
2. Cipher Text Policy ABE

In KP-ABE, a set of attributes is related to cipher text and the user's decryption key is linked with a tree structure. On the other hand, in CP-ABE, every user is connected with a set of attributes. The secret key is created based on the elements. In this case, threshold access structure is identified by the encrypted according to the concerned attributes. This message is then encrypted recognized on access structure such that only those whose qualities satisfy the access structure can decrypt it. With CP ABE technique, data is secure and confidential against attacks due to encryption.

In the proposed scheme, we have used KP-ABE scheme in which a number of keys are used for encryption and decryption purpose.

b) Advanced Encryption Standard (AES)

Many organizations have adopted AES across the world. It includes smart cards to big server applications; due to its simplicity AES are used. DES becomes obsolete with the increase of microprocessor chips and advancement in technology. So there was a need of new encryption technique. Rijndael's AES is an iterated block cipher with variable block and key lengths. The block length and the key length can vary and each of 128, 192, or 256 bits in size. Block and the intermediate cipher may also be estimated as a two-dimensional array of four rows referred to as the State. The length of columns varies relying on the bit length. In Rijndael, every operation is carried out either on single byte or on the four-byte word. Key can also be viewed same as this layout. Depending on the block measurement, the plain text offers as input to the cipher and is a one-dimensional array of sixteen, 24 or 32 bytes. In the state, these bytes are mapped in column order. AES is an iterative algorithm. A dissimilar key derived from the initial key is used in each of the iterations, called a round. The number of rounds depends on the key. Its operation is sub divided into 4 parts [21]. These are:

1. **ByteSub Transformation**

   This operation is a non-linear byte substitution. In this step, with the substitution field, every byte from the input state is changed by means of yet another byte. On each and every cell of the State, this transformation works independently.

2. **Shift Row Transformation**

All the rows participate in this transformation. By using the block length offsets of each row are determined. In case of Decryption, to neutralize the effect, the rows are shifted back, called InvShiftRow. That is, the rows are cyclically shifted left with an offset equal to a number of columns of State minus the offset for Encryption [22].

3. **Mix Column Transformation**

On every column of the State, this transformation works independently. A polynomial is the single column of the State on which this transformation works.

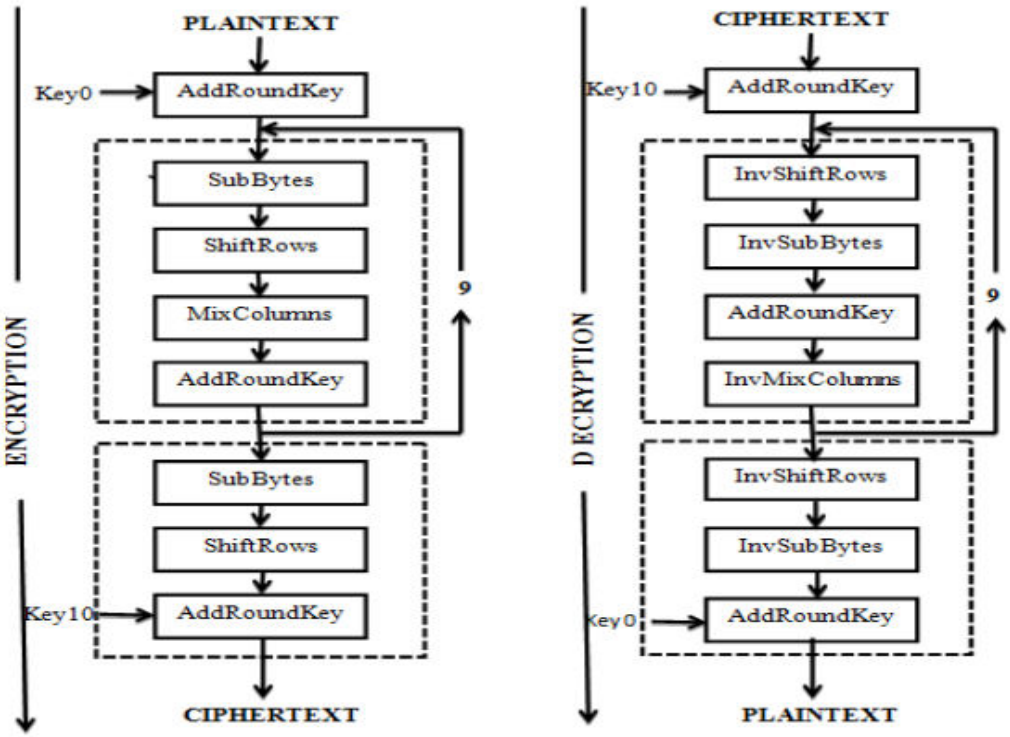

Figure 3.2: AES Encryption and Decryption

4. **Round Key Addition**

In this transformation, the circular key's combined with the State. This addition in GF (28) is a straightforward carried out with bit intelligent XOR. The circular key and State each have the same size. It is derived from the initial cipher by means of Key Schedule.

c) Exclusive OR (XOR)

XOR operation is widely used in cryptography and is also included as a part of standard encryption algorithms like AES and much more.

d) Blow Fish

Blowfish is an encryption algorithm which can be used as a replacement of DES algorithm. It is a symmetric algorithm which uses variable length keys from 32-bit keys to 448 bits and can be used for domestic and exportable use. It is calculated with 32-bit instruction processors in mind; it is meaningfully quicker than DES. Since its beginning, it has been evaluated significantly. Blowfish is unpatented, license-free, and available free for all uses.

e) RSA

RSA is a procedure used by modern processors to encode and decode messages. It is an asymmetric cryptographic algorithm. Asymmetric means that there are two dissimilar keys. This is also named as public key cryptography since one of them can be given to everybody. The other key must be kept private. It is established on the information that discovery the factors of an integer.

3.4 Attribute Based Encryption

There have been numerous of the schemes, planned for encryption. Such as, a simple encryption technique that is classically studied. ABE is better than other existing cryptography algorithms and provide better efficiency. The key intention for these models is to offer safety and access control. The predominant features are to deliver flexibility, scalability and fine grain to get admission to manipulate. In the classical model, this could be accomplished handiest whilst consumer and servers are dependent on area. But what if their domains are not trusted or no longer identical?

So, the brand new get admission to manipulate scheme this is ‘Attribute Based Encryption (ABE)’ scheme become delivered which encompass key policy characteristic primarily based encryption (KP-ABE). As compared with classical version, KP-ABE provided high-quality grained which grants right of entry to control.

In the cloud computing, for efficient revocation, get entries to manipulate mechanism primarily based on KP-ABE and an encryption method used together. It allows a data owner to lower-most of the computational overhead to the servers. The KP-ABE scheme presents best-grains to get right of entry to control. Each document or message is encrypted with a symmetric information encryption key (DEK), that is again encrypted via a public key, this is similar to a fixed of attributes in KP-ABE, that's generated corresponding to an get right of entry to tree shape. The encrypted facts report is stored with the corresponding attributes and the encrypted DEK. If the corresponding attributes of a document or message saved within the cloud satisfy to get entry to shape of a user's key, then the user is capable of decrypting the encrypted DEK. That may be used to decrypt the record or message.

CHAPTER 4

PROBLEM STATEMENT

4.1 Problem Statement

Cloud computing is an emerging field of research. Data security and privacy both are crucial issues that may become an obstacle in various cloud applications. Cloud service providers want to secure their infrastructure and software platforms from the hackers. The users ensure that provider has taken proper security actions to protect their data. Therefore security is needed on both the sides. In recent years, mass distributed storage became important to increase the data storage. Therefore, to access the sensitive data is one of the important issues of security in cloud computing. High level of performance is required to implement Mass Distributed Storage (MDS). It needs improvement in security in which threats come from any side. As a result, malicious attackers can attack easily in this network. Mainly attacks are launched during data transmission. It is impossible to balance security and functional concern; otherwise, the cost will become a big factor.

Currently, data is stored in the cloud which can be easily hacked by the hackers. They can access your personal data. In the proposed scheme, ABE based secure distributed storage scheme for big data in the cloud (ASDSS) has been proposed to provide security for the massive distributed data for big data in different clouds. In this, data is distributed into three parts (according to its sensitivity level) and all parts are encrypted using different techniques that include ABE, SA-EDS, XOR, and Blowfish. After encryption, these data parts will be stored in different cloud servers and then same keys will be used for retrieval and decryption of data. Hence, this technique provides efficient performance than SA-EDS, XOR, RSA, and Blowfish in terms of data storage and data retrieval.

In this work, we have considered three parameters for comparison purposes. These parameters are:

a) Data Storage

It has the capacity to store data. In the proposed scheme, proposed ASDSS technique takes less time to store in GB, MB, and KB respectively as compared to other existing techniques.

b) Data Retrieval

It is the time to access the store data. In the proposed scheme, proposed ASDSS technique takes less time to store in GB, MB, and KB respectively as compared to other existing techniques.

c) Total Time taken

It is the total time taken by the techniques for encryption and decryption both.

CloudSim tool is used for the implementation purpose. By the usage of CloudSim, developers can focus on unique systems design problems that they want to research, without getting concerned approximately info associated with cloud-based totally infrastructures and services. Experimental results show that this technique takes less time to store and retrieve data in KB, MB, and GB respectively. SA-EDS is compared with ASDSS and it is proved that ASDSS technique takes less time for data storage and data retrieval in KB, MB, and GB respectively.

4.2 Research Gaps

While big data resolves many difficulties such as high volumes of data, it is a constantly changing area that is always in growth and that still possess some issues. Cloud computing and big data security is a current topic or research. This problem becomes a matter to corporations when seeing uploading data onto the cloud. Questions such as who is the real owner of the data, where is the data, who has access to it and what kind of permissions they have, are hard to describe.

Some existing researches for the security of big data in the cloud are as follows:

1. Secure cloud computing for big data has been proposed for big data cloud. A hierarchal structure for big data analysis had been discussed. It had some security issues which were solved with help of identity-based solution.

2. To improve big data security, computing on mass data scheme had been proposed. It stores big data with the help of a number of techniques.
3. To improve the security of unstructured big data, sensitive level of data has been considered. Security suite was designed to recover unstructured big data. With the help of data classification, performance of the system had been enhanced.
4. In the existing work, data is stored on the same cloud for all types of data like financial, confidential and other. Cloud can be accessed by the hackers easily therefore, there is a need to improve the security of the cloud.

4.3 Objectives

Main objectives of the study are:

a) To study existing security issues and solutions for big data in the cloud.

b) To design and implement a novel technique for cloud security to enhance security for the massive data using XOR, AES, ABE.

c) To compare and analyze the performance of the proposed scheme with existing algorithms in terms of execution time and data storage.

CHAPTER 5

PROPOSED SCHEME

We implement the architecture in a way that it provides full security to its providers. Prior to storage, the input is divided into functional units. In this approach, the input is divided into three parts and then encrypted with different keys and stored in the cloud. In this work, data is divided into three categories according to its level of sensitivity. First is less sensitive, secondly is sensitive and the last one is more sensitive as shown in Figure 5.1. Different types of keys are applied over these different levels of sensitive data. To address the security, an enhancement for the mass distributed storage is used for the cloud security.

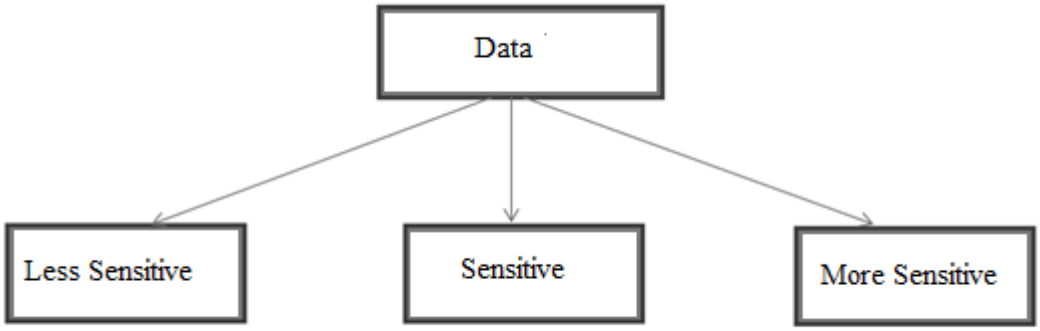

Figure 5.1: Data Division

5.1 ABE based Secure Distributed Storage Scheme for Big Data in Cloud (ASDSS)

This scheme is proposed to provide the security to the cloud storage at a high level. First of all, an input is divided into three categories [13]. One is less sensitive, secondly is sensitive and last is more sensitive. Therefore, the input is portioned into three levels of data. After that, a variety of keys is applied to different levels of sensitive data. The more powerful key is applied on the most sensitive data. Encryption is done in this step. For encryption purposes, XOR key is used for less sensitive data, AES is used for the sensitive and ABE is used for the more sensitive data. At the end, data is stored on different clouds. To decrypt the data, it is collected

from the different cloud storage and is decrypted with the keys. After this, merge the data to get the original one[14].

The two main features of attribute based encryption are given below:

1. Complex access control policies can be talked.

2. The exact list of operators need not be recognized properly. Knowledge of the access policy is adequate.

Attribute-based encryption (ABE) can be used for log encryption. Instead of encrypting each part of a log with the keys of all recipients, it's far possible to encrypt the log handiest with attributes which fit recipients' attributes. This primitive can also be used for broadcast encryption a good way to lower the number of keys used.

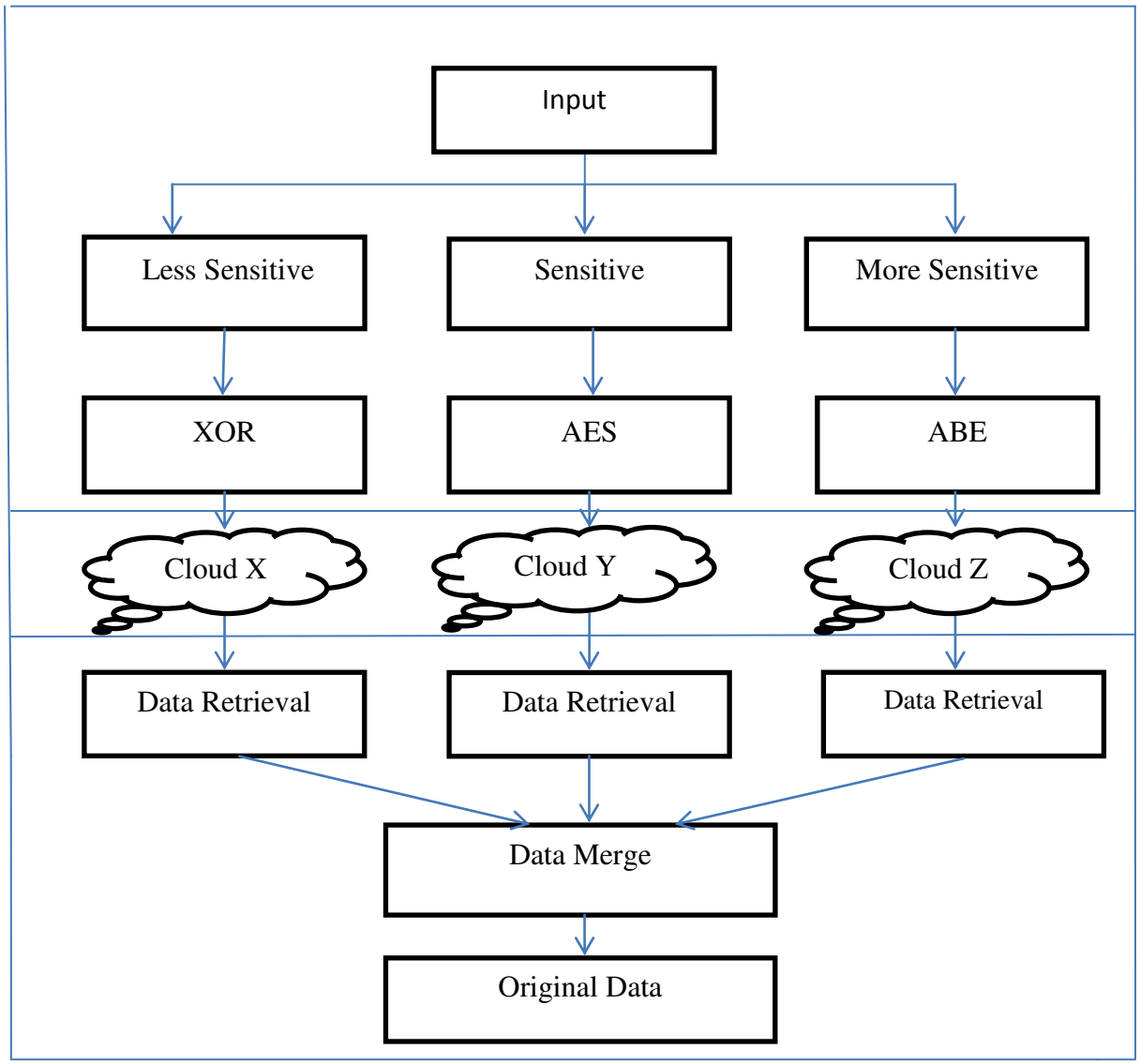

Figure 5.2: Flowchart of ASDSS

The entire process is divided into 3 phases as shown in Figure 5.2, A brief description of each phase is as follows:

Phase 1: In this phase, the input is taken. Data is entered as an input. According to the type of data, it is divided into less sensitive, sensitive and more sensitive data categories. Passwords and User id comes under more sensitive data. After this encryption is done using the keys. XOR key is used for less sensitive data, AES for sensitive data and ABE for more sensitive data. The whole process is done before the data is sent to the cloud. After this, cipher text is sent to the clouds.

Phase 2: In this phase, encrypted data is stored in the cloud. Three clouds are used for the proposed scheme for the storage of encrypted data. Cloud X, Cloud Y, and Cloud Z are used for the data storage on the cloud.

Phase 3: In this phase, encrypted data is retrieved from different clouds. Again, for decryption, same keys are applied on it. After that decryption data is merged and at the end, we get original and secure data.

a) Data Distribution and Encryption Algorithm

This algorithm is designed to divide the data into three parts[19] i.e. Less sensitive, Sensitive and More sensitive. The process of distributing the data is done by grouping the named-data-packets by using name labels. The inputs for this algorithm includes Named-data-packets (NDP)[13], Pre-stored Name List for more sensitive data(PNL 1), Pre-stored Name List for sensitive data(PNL 2). L_i are few name labels which each NDP has. The output of this algorithm includes different data packets after distribution according to their level of sensitivity.

The pseudo code of Data Distribution and Encryption Algorithm is shown in Algorithm 1. The main steps of this algorithm are described as follows:

Step 1: Input two Pre-stored Name List (PNL 1, PNL 2), one for More sensitive and other for Sensitive data packets respectively and searchable named-data-packets (NDP).

Step 2: For all NDP, we will search each data packet and check whether it belongs to PNL 1 or PNL 2 or none of these.

Step 3: If the match is found in PNL 1, ABE algorithm will be executed to encrypt the data.

Step 4: If a match is found in PNL 2, AES algorithm will be executed to encrypt the data.

Step 5: Otherwise, XOR operation will be executed to encrypt the data packets.

Step 6: Output all the encrypted data packets, that includes α, β, γ and then store these encrypted data packets separately in different cloud servers.

Algorithm 1 Data Distribution & Encryption Algorithm

Require: NDP, PNL1, PNL2

Ensure: D, α, β, γ

1.Input NDP, PNL1, PNL2

2.READ: Read data

3.**for** ∀ NDP **do**

4.**for** each data packet **do**

5.if ∃ a $L_i \in$ PNL 1 **then**

6.Key K1 is generated using KeyGen(PK, A, MK)

7.Execute ABE Algorithm to encrypt the data using key K1

8.Generate α

9.**else if** ∃ a $L_i \in$ PNL 2 **then**

10.Key K2 is generated using KeyGen(PK, A, MK)

11.Execute AES Algorithm to encrypt the data using key K2

12.Generate β

13.**else**

14.Randomly generate a key K3

15.Do XOR operation to encrypt the data with key K3

16.Generate γ

17.**end if**

18.**end for**

19.Obtain the values of D

20.**end for**

21.Output α, β, γ

b) Data Retrieval Algorithm

This algorithm is designed to retrieve the original data that was first distributed as shown in Algorithm 1. The inputs that are included in this algorithm are α, β, γ, K1, K2, K3. The output of this algorithm will be the original data D [17].

The pseudo-code of this algorithm is shown in Algorithm 2. The main steps in Data Retrieval Algorithm are described as follows:

Step 1: In this step, we will input the encrypted data packets that we had obtained from Algorithm 1 and keys (stored in special register[13]) will be needed to decrypt the data respectively.

Step 2: Then we will initialize few datasets ʎ, ʎ', ʎ" which are used to store the data after decryption.

Step 3: Then we will decrypt data from different cloud servers using keys and algorithms respectively.

Step 4: After obtaining the decrypted data, we will combine these decrypted data packets to obtain the original data.

Step 5: Output the original data

Algorithm 2 Data Retrieval Algorithm
Require: α, β, γ, K1, K2, K3
Ensure: D
1. Input α, β, γ, K1, K2, K3
2. Initialize ʎ←0, ʎ' ←0, ʎ'' ←0
3. /* User receives inputs α, β, γ from different cloud servers*/
4. ʎ ← Decrypt α with key K1 using ABE algorithm
5. ʎ' ← Decrypt β with key K2 using AES algorithm
6. ʎ'' ← γ ⊕K3
7. D ← Combine ʎ, ʎ' and ʎ'' to obtain original data
8. Output D

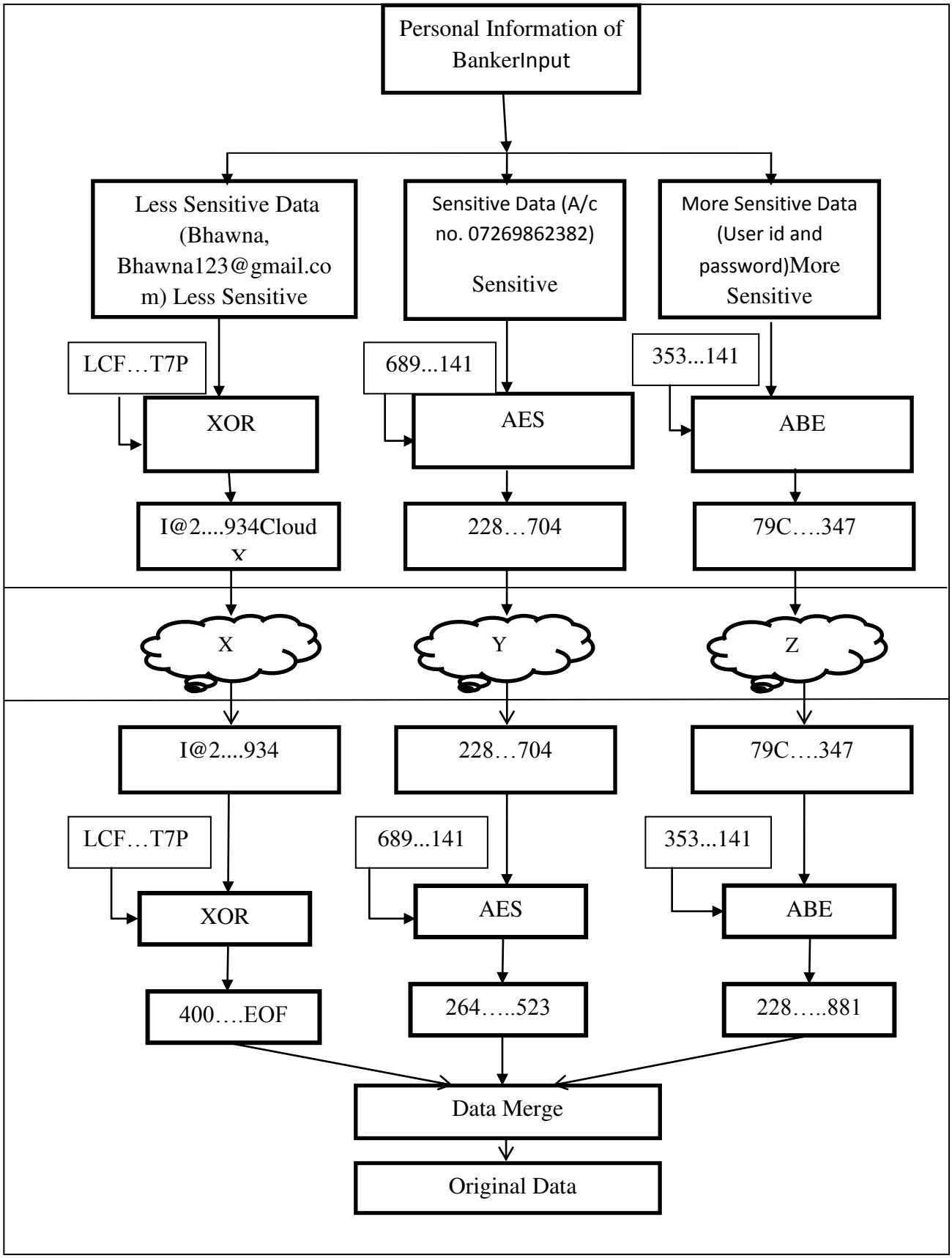

Figure 5.3: ASDSS with actual data set

As we discussed above, we are working to enhance the security of the network. ABE, AES and XOR keys are applied to get optimal output and enhance the security of the cloud. We are using 3 keys for the data encryption and similarly, 3 clouds are used for data storage as shown in Figure 5.3. ABE, AES, and XOR are used for this whole process.

5.1.1 PSEUDOCODE OF ABE ALGORITHM

ABE operation is applied with key K1 on more sensitive data. As a result, we get α as an output. It is a least timing consuming process.

Pseudocode

```
Step 1: REQUIRE: Alpha, K1, c1
Step 2: READ DATA: More sensitive data
Step 3: SETUP KEYS PARAMETERS: pK, mK, sK
Step 4: Key generation: initialize attributes, pK, mK, sF
then: serialize key (k1)
Step 5: PERFORM ENCRYPTION: pass (encryption_file, policy,
PK, output_file)
then: serialize text into ciphertext
Step 6: PERFORM DECRYPTION: pass (cipherText, pK)
then: unserialize (c1, pK, sK, pK)
Step 7: OUTPUT: alpha
```

5.1.2 PSEUDOCODE OF AES ALGORITHM

AES is another important algorithm which can be applied for security purpose. In our work, we applied it to sensitive data so that it can be protected from hackers. The K2 key is used for encryption and decryption.

Pseudocode

```
Step 1: REQUIRE: beta, K2, c2
Step 2: INITIALIZE PARAMETERS : SEC_LEVEL, MAX_ITERS
Step 3: READ DATA: read data in bytes
Step 4: SETUP KEYS PARAMETERS: sK, N
Step 5: INITIALIZE KEY GENERATORS: set(SEC_LEVEL)
  then: serialize key (k2)
Step 6: PERFORM ENCRYPTION: pass (plaintext) then
   serialize plaintext into cipher text
```

```
Step 7: PERFORM DECRYPTION: read(bytesFromFile, k)
  then: unserialize (c2, k)
Step 8: OUTPUT: beta
```

5.1.3 PSEUDOCODE OF XOR ALGORITHM

XOR operation is applied with Key K3 on the less sensitive data. As a result, we get as an output. It is the most time-consuming process.

Pseudocode

```
Step 1:Initialize TextArray to GetData from Massive_Data
Step 2: Initialize Encrypt_Array to NULL
Step 3: Start Time_Var to ecript the data
Step 4: PERFORM FOR Loop until the String(TextArray) is
not equal to Null;
Step 5: START Encryption on String(TextArray)
using XOR( )Implementation;
Step 6: GENERATE Random Keys by using Text_Array
Step 7: PERFORM XOR( ) Implementation on Text_Array
with
using String and Key;
Step 8: INCREMENT Time_Var;
Step 9: RETURN to Decryption(encrypted_data,Keys);
Step 10: PERFORM For Loop UNTIL the String(TextArray) is
not equal to Null;
Step 11: PERFORM Decryption on Data getting from
Encryption by using its Keys
Step 12: CALCULATE TOTAL Time
```

Table 5.1: Key Length of various Methods

Techniques	Key Length (in Bits)
XOR	10
ABE	48
AES	64
Blowfish	32
RSA	64

CHAPTER 6

RESULTS AND DISCUSSIONS

6.1 Implementation

CloudSim presents a generalized and extensible simulation framework that permits seamless modeling and simulation of the app in overall performance. By the usage of CloudSim, developers can focus on unique system design problems that they want to research, without getting concerned approximately info associated with cloud-based totally infrastructures and services. CloudSim, that's a toolkit for the modeling and simulation of Cloud computing environments involves the rescue as shown in Figure 6.1. It presents machine and behavioral modeling of the Cloud computing additives. Simulation of cloud environments and applications to assess overall performance can provide useful insights to explore such dynamic, massively allotted and scalable environments.

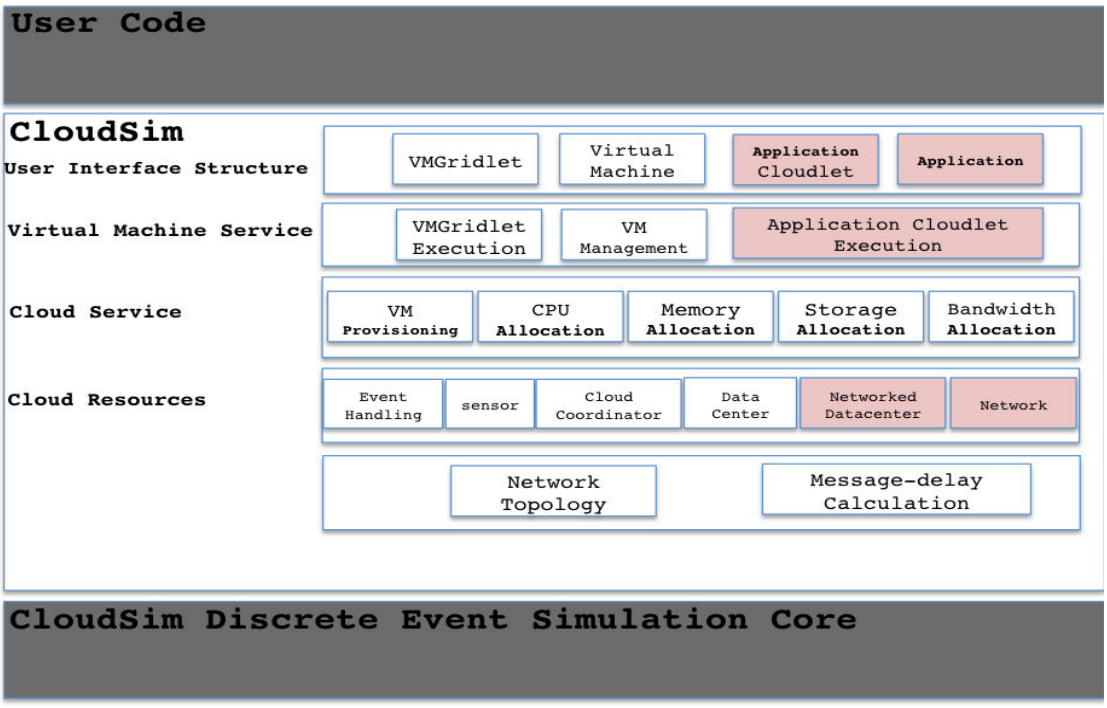

Figure.6.1: CloudSim Architecture

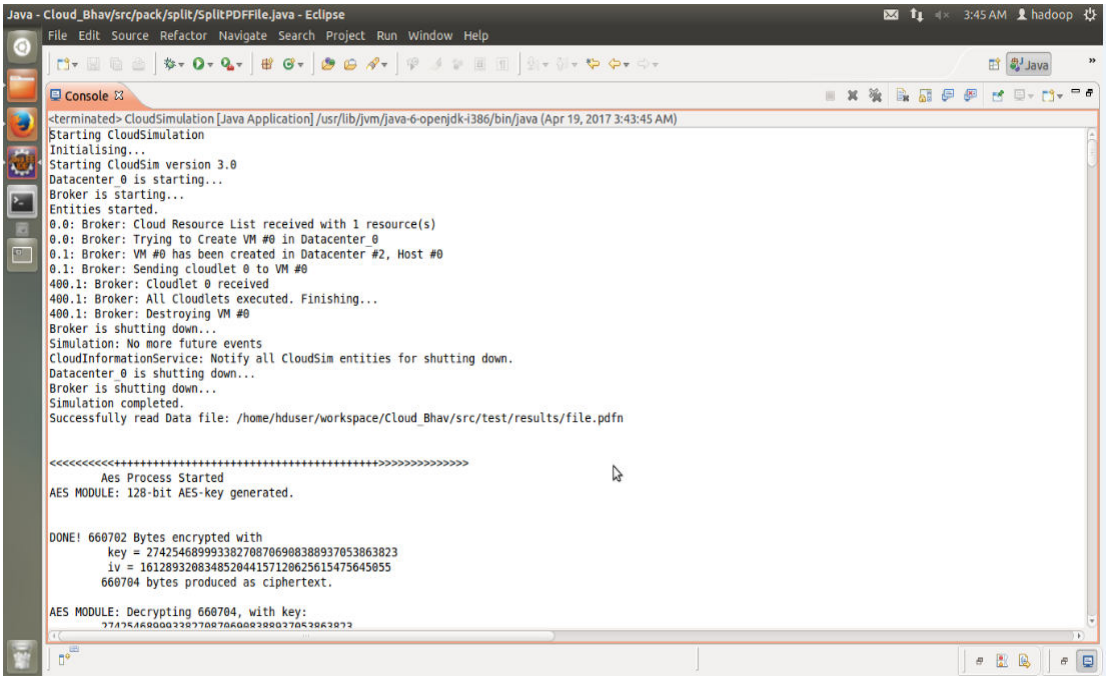

Figure 6.2: Initialization of AES Process

Figure 6.2 depicts the initialization and starting process of AES scheme. Keys for both encryption and decryption are shown with the cipher text. Similarly, keys are generated for all the rounds of AES.

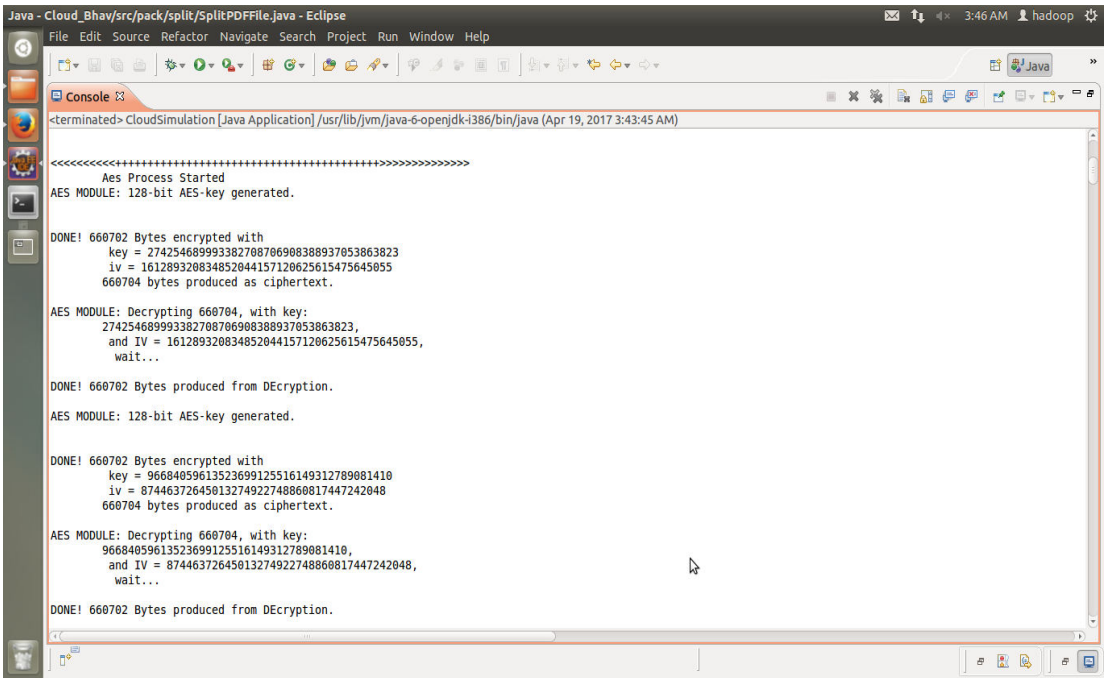

Figure 6.3: AES Process

Figure 6.3 depicts 128-bit AES-key generation. Keys for both encryption and decryption are shown with the cipher text. Keys are generated for all the rounds of AES.

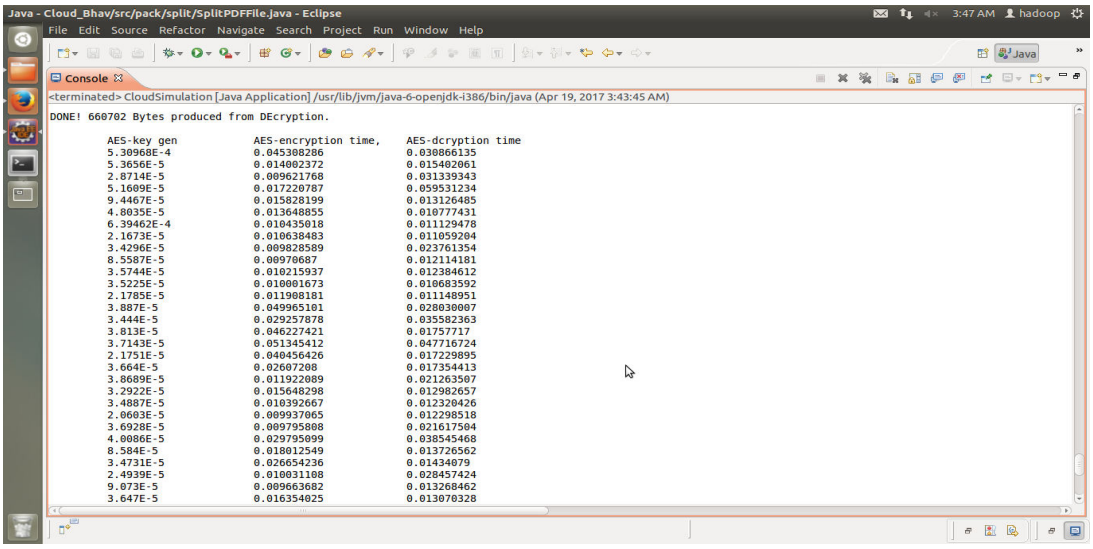

Figure 6.4: AES Key Generation

Figure 6.4 depicts AES key generation. It also shows time for AES encryption and AES Decryption.

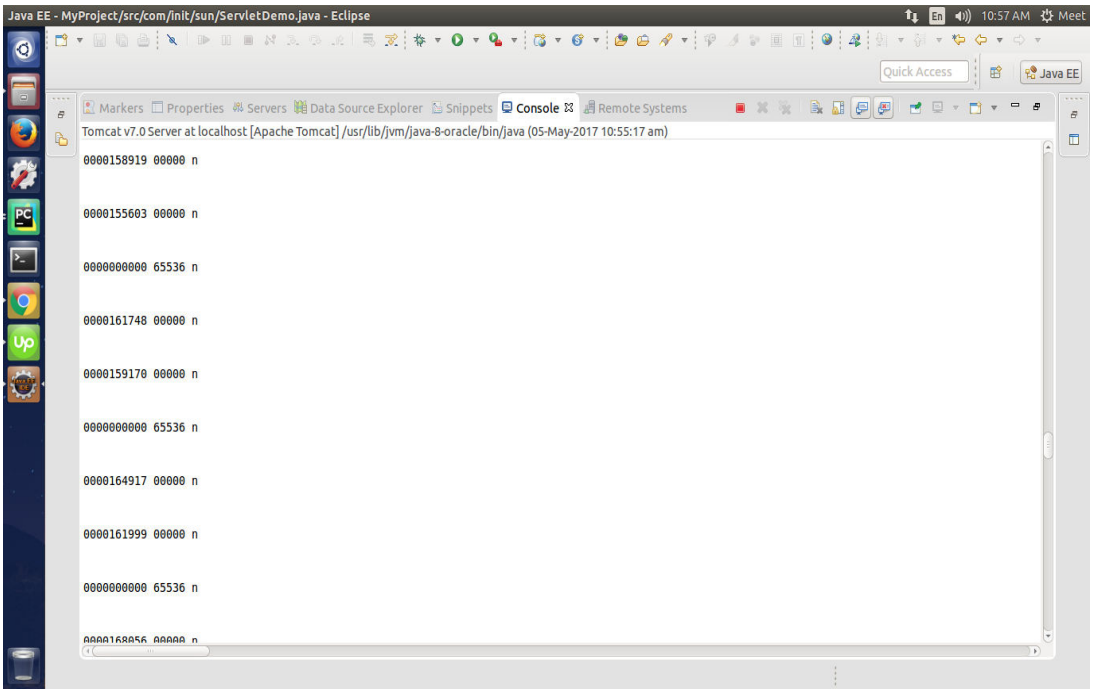

Figure 6.5: File uploading process for ABE

Figure 6.5 depicts next step of file uploading process for ABE. After this process, the key is generated to encrypt the data. As a result, we get encrypted data.

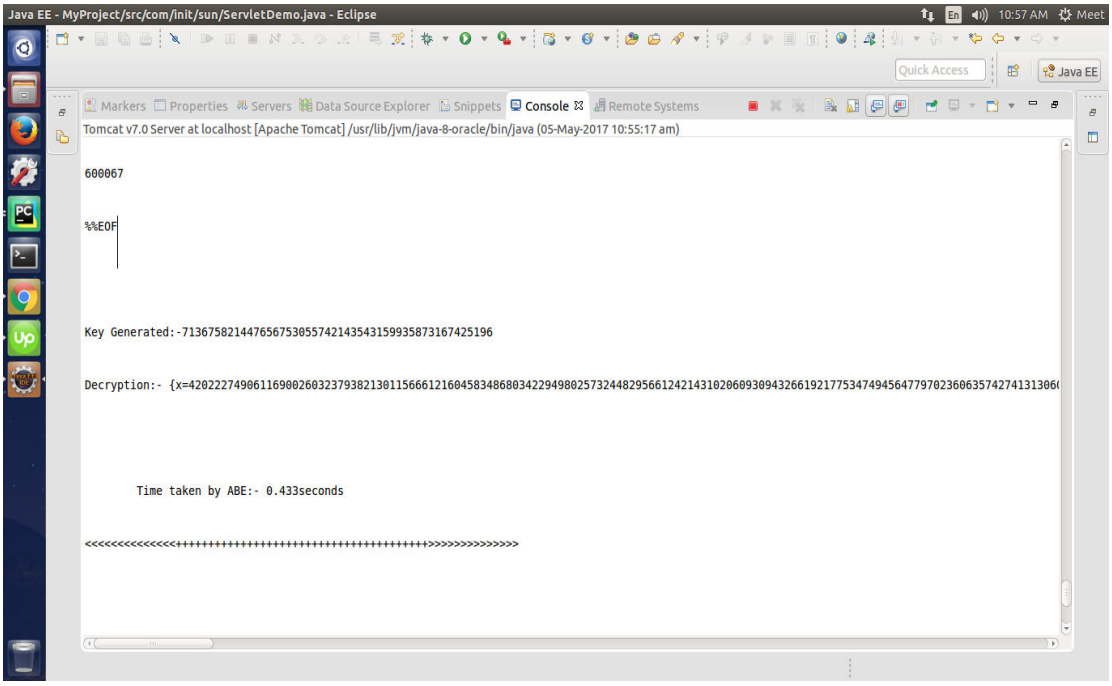

Figure 6.6: Key generated for ABE

Figure 6.6 depicts generated a key for the encryption process. It also shows decrypted data and total time taken by ABE for the whole process.

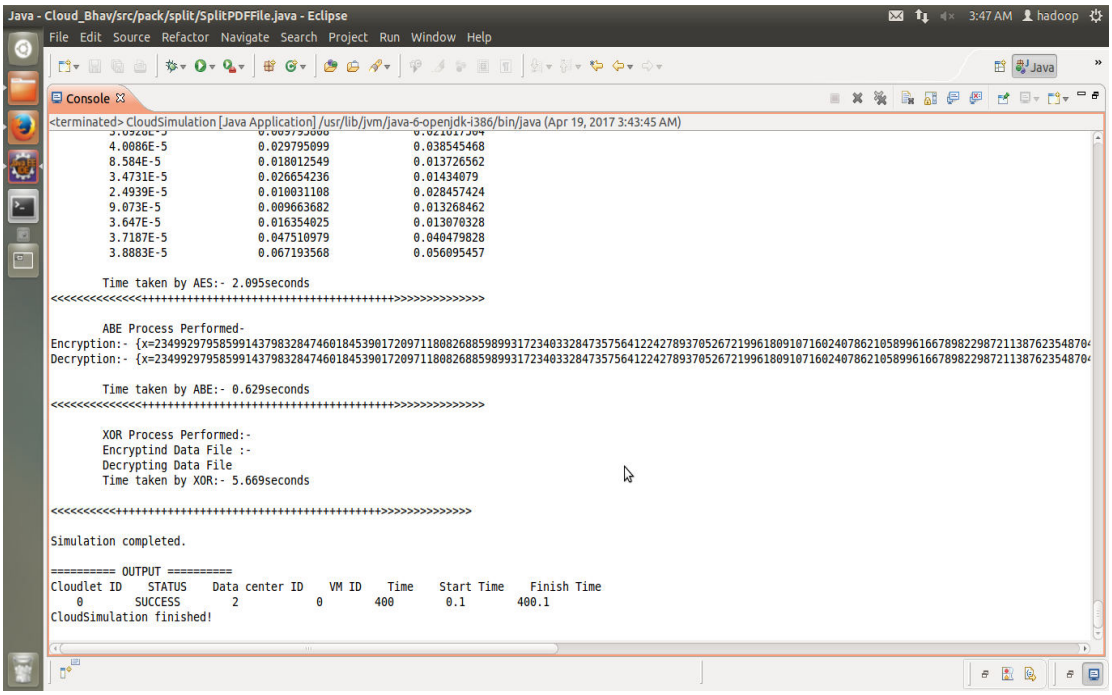

Figure 6.7: ABE and XOR Processing Time

Figure 6.7 depicts total time taken by AES scheme. Further, it also shows time taken by ABE and XOR schemes. It is concluded that ABE takes less time i.e. 0.629 seconds, XOR takes 5.6669 seconds and AES take 2.095 seconds.

6.2 Simulation Results and Discussions

In the proposed scheme, ASDSS technique is implemented. We have used CloudSim Toolkit 3.0.3, Eclipse as IDE and Java 1.8 as a platform to implement this scheme. In this, an example of financial data is taken as an input that is to be encrypted and stored in the cloud. Firstly, Algorithm 1 will get executed for distribution, encryption of data and distribution of financial data for implementation is shown in TABLE 1 and finally, Algorithm 2 will get executed for data retrieval. Experimental results show that this technique takes less time to store and retrieve data in KB, MB, and GB respectively. SA-EDS is compared with ASDSS which proved that ASDSS technique takes less time for data storage and data retrieval in KB, MB, and GB respectively.

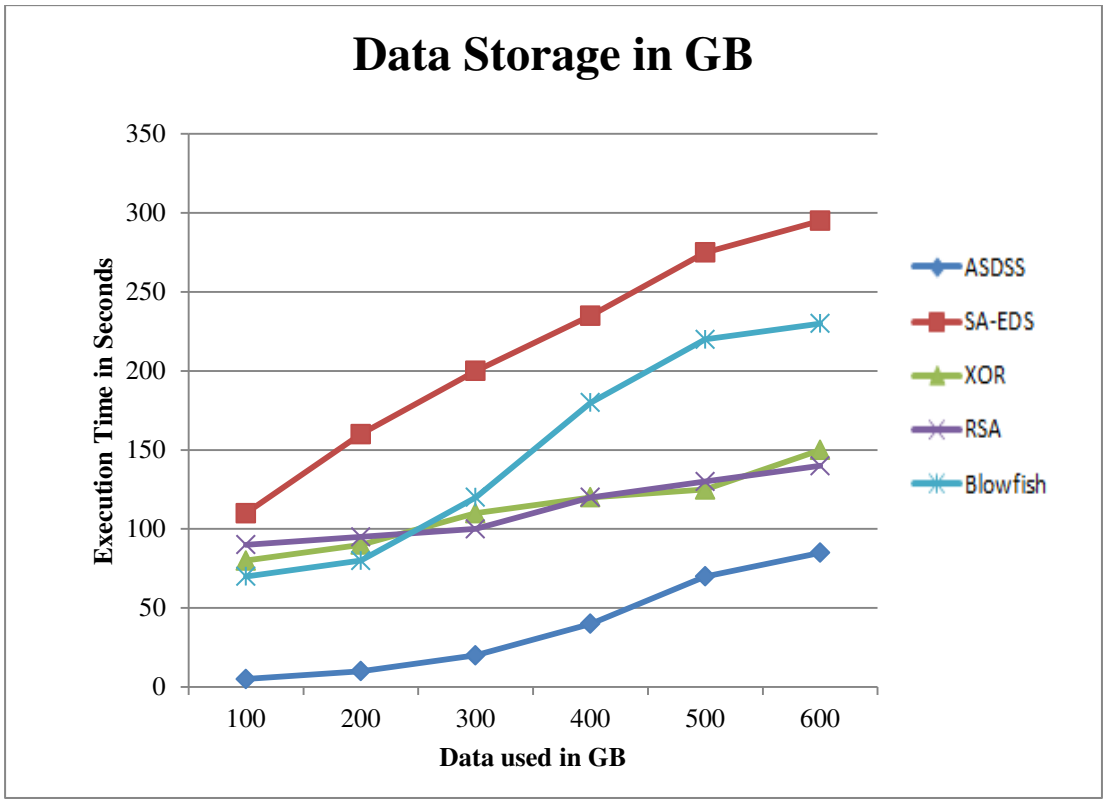

Figure 6.8**:** Comparison between data storage for ASDSS, SAEDS, XOR, Blowfish, and RSA in GB

Figure 6.8 depicts that ASDSS takes less time for storage than all other methods. So, ASDSS is more efficient technique.

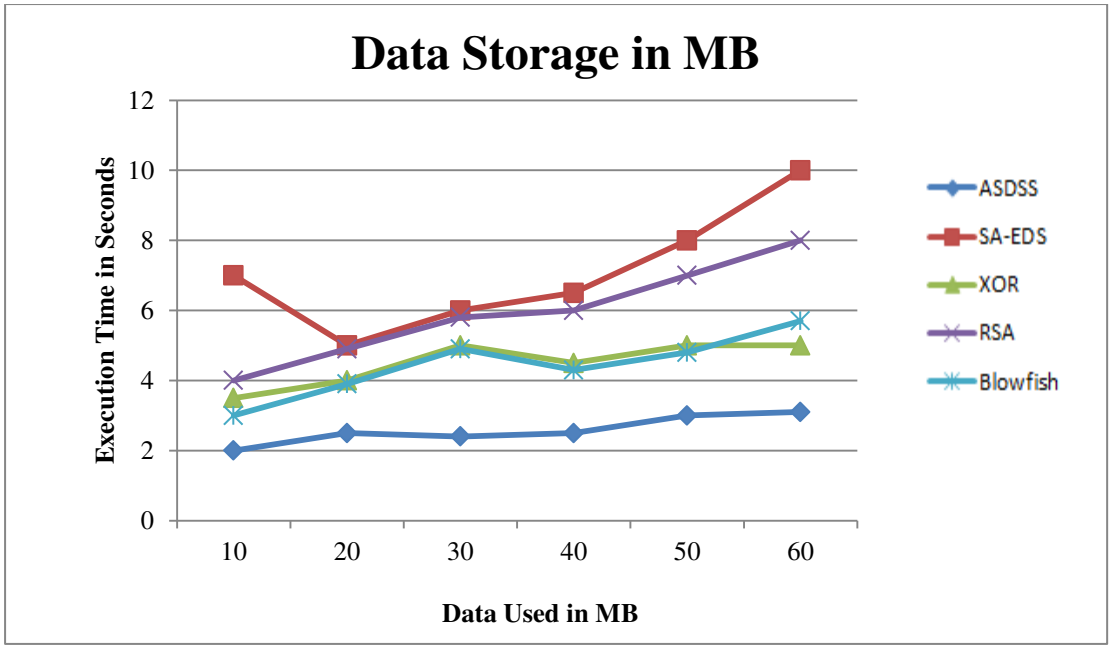

Figure 6.9: Comparison between data storage for ASDSS and other methods in MB

Figure 6.9 depicts a comparison between data storage in MB for ASDSS and other methods. It shows that the ASDSS takes less time for data storage.

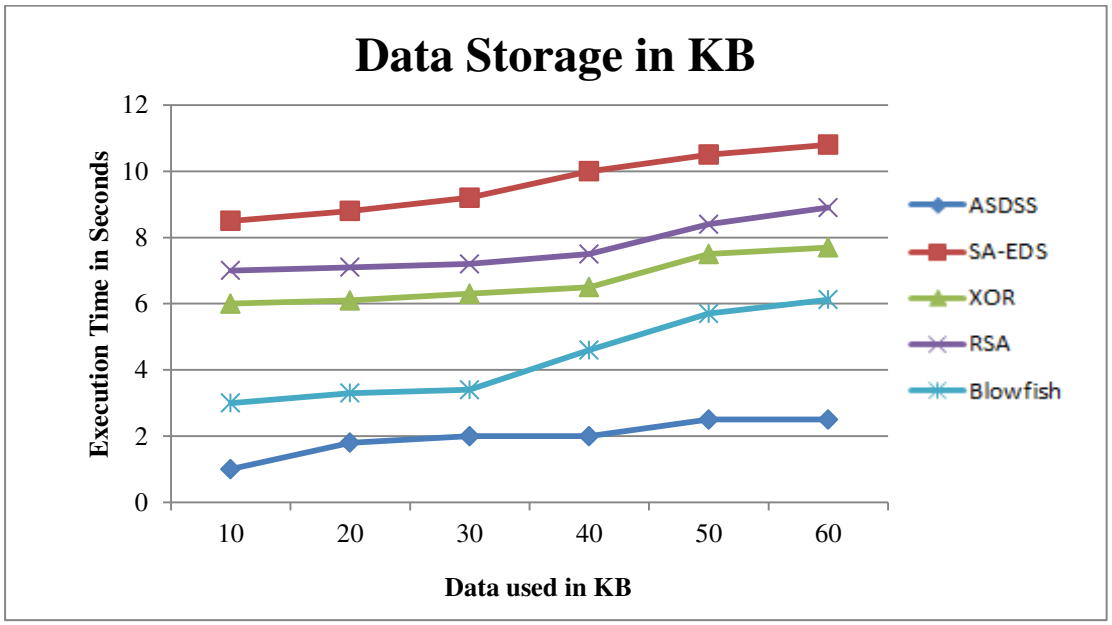

Figure 6.10: Comparison between data storage for ASDSS and other methods in KB

Figure 6.10 depicts the comparison between ASDSS and other methods. It is proved that ASDSS provides better performance as compared to AES, Blowfish, RSA, and XOR, as it takes less time for data storage in KB.

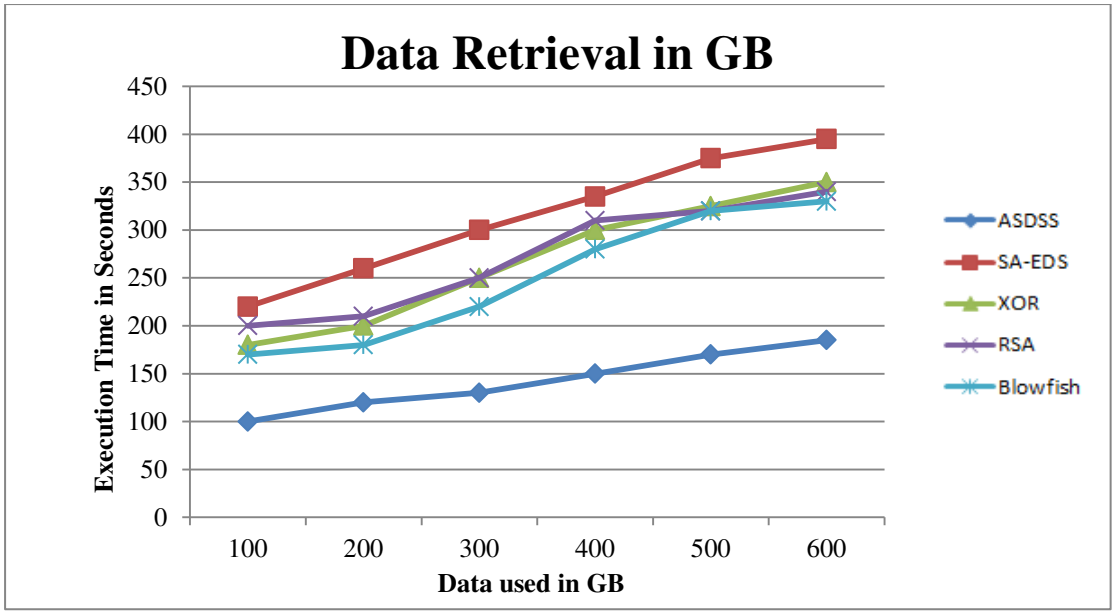

Figure 6.11: Comparison between data retrieval for ASDSS and other methods in GB

Figure 6.11 depicts the comparison between ASDSS and other methods. It is proved that ASDSS provides better performance as compared to AES, Blowfish, RSA, and XOR, as it takes less time for data retrieval in GB.

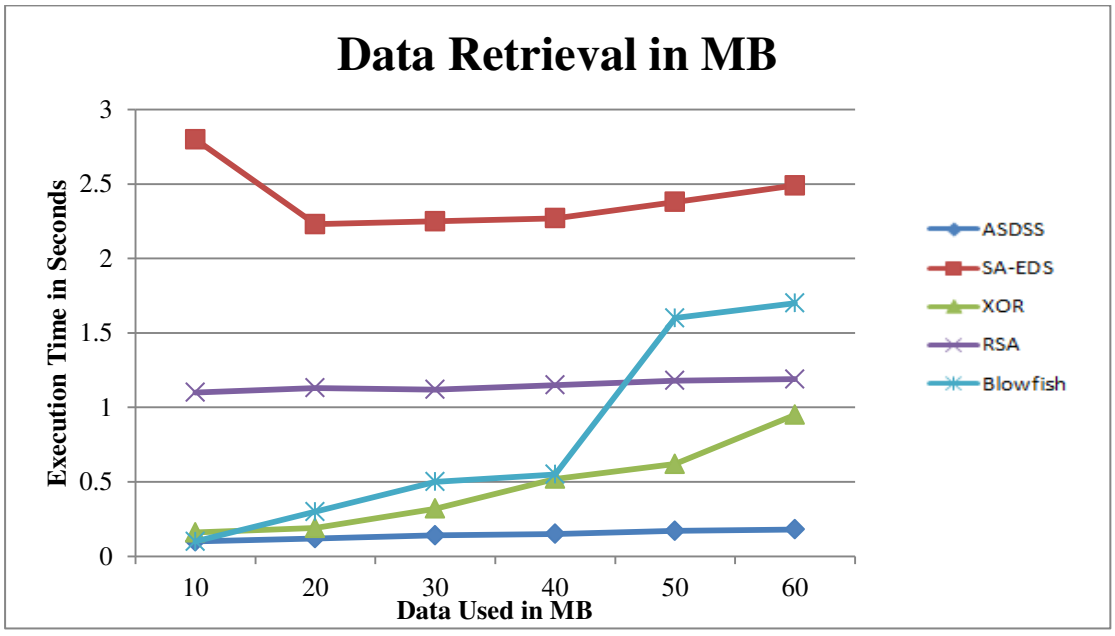

Figure 6.12: Comparison between data retrieval for ASDSS and other methods in MB

Figure 6.12 depicts the comparison between ASDSS and other methods. It is proved that ASDSS provides better performance as compared to AES, Blowfish, RSA, and

XOR, as it takes less time for data retrieval in MB but it is more than data retrieval in KB.

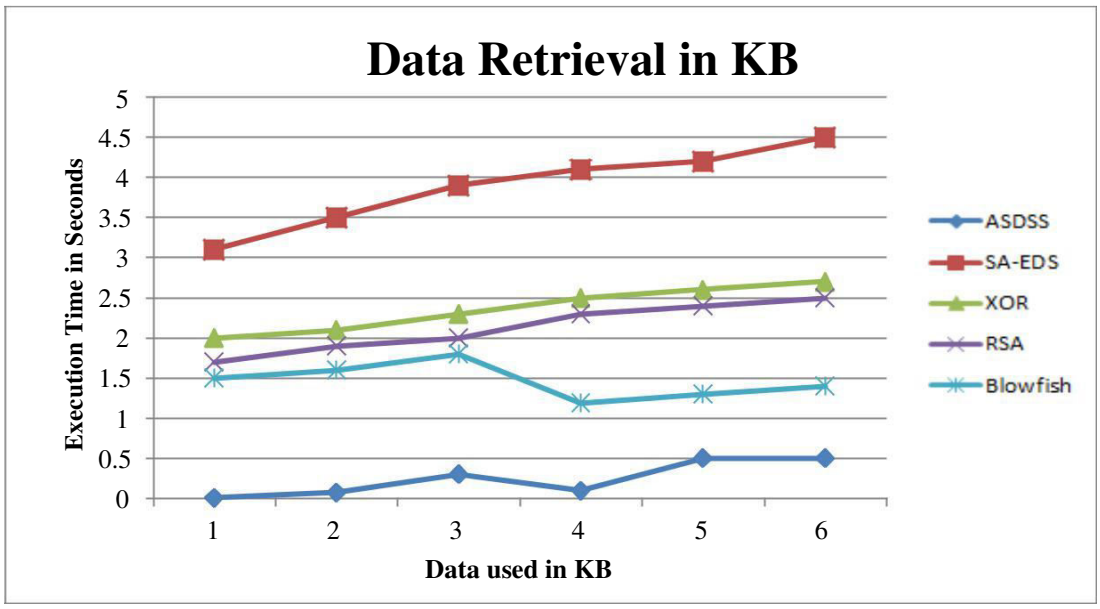

Figure 6.13: Comparison between data retrieval for ASDSS and other methods in KB

Figure 6.13 depicts the comparison between ASDSS and other methods. It is proved that ASDSS provides better performance as compared to AES, Blowfish, RSA, and XOR, as it takes less time for data retrieval in KB.

CHAPTER 7

CONCLUSION AND FUTURE SCOPE

7.1 Conclusion

Over the past three decades, many different methods have been explored by a large number of scientists for security purposes. A variety of approaches has been proposed and tested by researchers in different parts of the world using various cryptography techniques and encryption algorithms. No algorithm in this world is 100% secure till date. Cloud computing provides a wide network access, storage space, computation resources, user and corporate applications. Cloud computing has many amazing features like on-demand user, pay-as-per-use, shared pool computations, rapid elasticity, on-demand services etc. Cost saving is the main advantage of the cloud. Security of cloud is the major disadvantage of the cloud. Information stored on the cloud server is the biggest challenge. The user is not comfortable to handle data in the cloud. To make the server secure, many encryption techniques have been developed using passwords. Massive Data Storage is a technique used in cloud computing to deal with the high storage, instead of security issues that include availability, accessibility, and reliability. Integration of system becomes complicated when the data size increases. The main concern of MDS is related to the security. Due to the restriction of computer resources, data synchronization problem occurs for big storage only. In the existing scheme, data is stored on the same cloud and only one key (XOR) had been used for encryption and decryption purpose. Data can be easily hacked in this scheme. So to overcome this problem, ASDSS scheme is proposed. In this, data is distributed into three parts (according to its sensitivity level) and all parts are encrypted using different techniques that include ABE, SA-EDS, XOR, and Blowfish. After encryption, these data parts will be stored in different cloud servers and then same keys will be used for retrieval and decryption of data. Hence, this technique provides efficient performance than SA-EDS, XOR, RSA, and Blowfish in terms of data storage and data retrieval. It is also proved that ASDSS takes less time to retrieve data in GB, MB and KB as compared to SA-EDS, XOR, RSA and Blowfish techniques.

7.2 Future Scope

We will enhance the security techniques for data storage and retrieval in order to improve the performance of the scheme and implement it in real time scenario. Future work will also address securing data duplications in order to increase the level of data availability since any data center's down will cause the failure of data retrievals.

REFERENCES

[1] S. F. Altschul, W. Gish, W. Miller, E. W. Myers, and D. J. Lipman, **"**Basic local alignment search tool,**"** *Journal of Molecular Biology*, 1990.

[2] Tene, Omer, and Jules Polonetsky, "Privacy in the age of big data: a time for big decisions", Stanford Law Review Online 64: 63, 2012.

[3] Majhi, Santosh Kumar, and Gyanaranjan Shial, "Challenges in Big Data Cloud Computing and Future Research Prospects: A Review**"**, *Smart CR* 5.4, pp: 340-345, 2015.

[4] Tallon, Paul P,"Corporate governance of big data: Perspectives on value, risk, and cost." *Computer* 46.6, pp, 32-38, 2013.

[5] Chang, Victor, "Towards a Big Data system disaster recovery in a Private Cloud**"**, *Ad Hoc Networks* 35, pp: 65-82, 2015.

[6] Snell, Addison,"Solving Big Data Problems with Private Cloud Storage**"**, *Intersect360 Research,* October 2011.

[7] A. Alahmadi, M. Abdelhakim, J. Ren, T. Li, **"**Defense against primary user emulation attacks in cognitive radio networks using advanced encryption standard**"**, *IEEE Trans. Inf. Forensics Secur*ity, Vol. 9, No.6, pp. 772–781.

[8] M. Ali, S. Khan, A. Vasilakos, **"** Security in cloud computing: Opportunities and challenges**"**, *Inf. Sci*. 305, pp.357–383, 2015.

[9] R. Aliev, W. Pedrycz, B. Fazlollahi, O. Huseynov, A. Alizadeh, B. Guirimov, **"**Fuzzy logic-based generalized decision theory with imperfect information,**"** Inf. Sci. 189, pp.18–42, 2012.

[10] G. Ateniese, K. Fu, M. Green, S. Hohenberger, **"**Improved proxy re-encryption schemes with applications to secure distributed storage,**"** *ACM Trans. Inf. Syst. Secur.,* Vol.9, No.1, pp.1–30, 2005.

[11] J. Baek, Q. Vu, K. Liu, X. Huang, Y. Xiang, **"**A secure cloud computing based framework for big data information management of smart grid*,", IEEE Trans. Cloud Computing*, Vol 3, No.2, pp. 233–244, 2015.

[12] Z. Brakerski, C. Gentry, V. Vaikuntanathan, **"**fully homomorphic encryption without bootstrapping,**"** *ACM Trans. Comput. Theory*, Vol 6, No.3, 2014.

[13] Li, Yibin, et al. "Intelligent cryptography approach for secure distributed big data storage in cloud computing." *Information Sciences*, 2016.

[14] G. Kalpana, P.V. Kumar, R.V. Krishnaiah, "Secured Cloud Computing Using User Classification and Bilinear Diffie-Hellman Schema", *Advanced Computing (IACC) IEEE 6th International Conference on*, pp. 563-568, 2016.

[15] Li, Yibin, "Intelligent cryptography approach for secure distributed big data storage in cloud computing." *Information Sciences*, 2016.

[16] A. Alahmadi, M. Abdelhakim, J. Ren, T. Li, **"**Defense against primary user emulation attacks in cognitive radio networks using advanced encryption standard,**"** *IEEE Trans. Inf. Forensics Secur.* Vol 9, No.5, pp.772–781, 2015.

[17] M. Ali, S. Khan, A. Vasilakos, **"**Security in cloud computing: Opportunities and challenges,**"** *Inf. Sci.*, pp.357–383, 2015.

[18] K. Gai, M. Qiu, L. Tao, Y. Zhu, "Intrusion detection techniques for mobile cloud computing in heterogeneous 5G", *Secure. Commun. Netw.*, pp.1–10, 2015.

[19] K. Gai, M. Qiu, B. Thuraisingham, L. Tao, **"**Proactive attribute-based secure data schema for mobile cloud in financial industry,**"**, *The IEEE International Symposium on Big Data Security on Cloud*, IEEE 17th International Conference on High Performance Computing and Communications, New York, USA, pp. 1332–1337, 2015.

[20] Li, Yibin, "Intelligent cryptography approach for secure distributed big data storage in cloud computing." *Information Sciences*, 2016.

[21] Abhilasha Naidu, A.Y. Deshmukh, Vipin Bhure **"**Design of High Throughput and Area Efficient Advanced Encryption System Core**"**, International Conference on Communication and Signal Processing, *IEEE, April 2014.International Conference on Communication and Signal Processing, IEEE,* April 2014.

[22] Bala, Tannu, and Yogesh Kumar, "Comparative Analysis of Parallel AES Algorithm with Pipelined AES Algorithm.**"**

[23] Li, Yibin, "Intelligent cryptography approach for secure distributed big data storage in cloud computing." *Information Sciences*, 2016.

[24] Zhao, Jiaqi, "A security framework in G-Hadoop for big data computing across distributed Cloud data centers." *Journal of Computer and System Sciences,* Vol.80, No.5, 2015.

[25] Manogaran, Gunasekaran, Chandu Thota, and M. Vijay Kumar, "MetaCloudDataStorage architecture for Big Data security in cloud computing." *Procedia Computer Science* 87, pp. 128-133, 2016.

[26] Liu, Chang, "External integrity verification for outsourced big data in cloud and IoT: A big picture**"**, *Future Generation Computer Systems* 49, pp-58-67, 2015.

[27] Ramachandran, Muthu, and Victor Chang. "Towards performance evaluation of cloud service providers for cloud data security." *International Journal of Information Management, Vol 6, No.4,* pp. 618-625, 2015.

[28] Baek, Joonsang, "A secure cloud computing based framework for big data information management of the smart grid." *IEEE transactions on cloud computing, Vol 3, No.2*, pp.233-244, 2015.

[29] Sookhak, Mehdi,"Dynamic remote data auditing for securing big data storage in cloud computing." *Information Sciences*, pp.101-116, 2016.

[30] Gai, Keke,"Privacy-aware adaptive data encryption strategy of big data in cloud computing." *Cyber Security and Cloud Computing (CSCloud), 2016 IEEE 3rd International Conference on* . IEEE, 2016.

[31] Pasupuleti, Syam Kumar, Subramanian Ramalingam, and Rajkumar Buyya. "An efficient and secure privacy-preserving approach for outsourced data of resource constrained mobile devices in cloud computing." *Journal of Network and Computer Applications* 64, pp.12-22, 2016.

[32] Sood, Sandeep K. "A combined approach to ensure data security in cloud computing." *Journal of Network and Computer Applications* 35, pp. 1831-1838, 2012.

[33] Shaikh, Rizwana, and M. Sasikumar. "Data classification for achieving security in cloud computing." *Procedia computer science* 45, pp.493-498, 2015.

[34] Li, Yibin, "Intercrossed Access Controls for Secure Financial Services on Multimedia Big Data in Cloud Systems." *ACM Transactions on Multimedia Computing, Communications, and Applications*, 2016.

[35] Qiu, Meikang, **"**Proactive user-centric secure data scheme using attribute-based semantic access controls for mobile clouds in financial industry**"**, *Future Generation Computer Systems*, 2016.

[36] Chandrasekaran, Balaji, and Ramadoss Balakrishnan. "Attribute Based Encryption Using Quadratic Residue for the Big Data in Cloud Environment." *Proceedings of the International Conference on Informatics and Analytics*. ACM, 2016.

[37] Kang, Seungmin, Bharadwaj Veeravalli, and Khin Mi Mi Aung. **"**A Security-Aware Data Placement Mechanism for Big Data Cloud Storage Systems." *Big Data Security on Cloud (BigDataSecurity), IEEE International Conference on High Performance and Smart Computing (HPSC), and IEEE International Conference on Intelligent Data and Security (IDS), 2016 IEEE 2nd International Conference on*. IEEE, 2016.

[38] Gai, Keke, Meikang Qiu, and Sam Adam Elnagdy. "Security-aware information classifications using supervised learning for cloud-based cyber risk management in financial big data." *Big Data Security on Cloud (BigDataSecurity), IEEE International Conference on High Performance and Smart*

Computing (HPSC), and IEEE International Conference on Intelligent Data and Security (IDS), 2016 IEEE 2nd International Conference on. IEEE, 2016.

[39] Gai, Keke, Meikang Qiu, and Sam Adam Elnagdy., "A novel secure big data cyber incident analytics framework for cloud-based cyber security insurance", *Big Data Security on Cloud (BigDataSecurity), IEEE International Conference on High Performance and Smart Computing (HPSC), and IEEE International Conference on Intelligent Data and Security (IDS), 2016 IEEE 2nd International Conference on*. IEEE, 2016.

[40] Chen, Deyan, and Hong Zhao, **"**Data security and privacy protection issues in cloud computing." *Computer Science and Electronics Engineering (ICCSEE), 2012 International Conference on*. Vol. 1. IEEE, 2012.

[41] Baker, T., "GreeDi: An energy efficient routing algorithm for big data on the cloud." *Ad Hoc Networks* 35, pp.83-96, 2015.

[42] Fazio, Maria, "Big data storage in the cloud for smart environment monitoring." *Procedia Computer Science* 52, pp. 500-506, 2015.

[43] Demirkan, Haluk, and Dursun Delen. "Leveraging the capabilities of service-oriented decision support systems: Putting analytics and big data in the cloud." *Decision Support Systems* 55.1, pp.412-421, 2013.

[44] Teli, Prasad, Manoj V. Thomas, and K. Chandrasekaran. "Big Data Migration between Data Centers in Online Cloud Environment." *Procedia Technology* 24, pp.1558-1565, 2016.

[45] Vennila, V, and A. Rajiv Kannan. "Symmetric Matrix-based Predictive Classifier for Big Data computation and information sharing in Cloud." *Computers & Electrical Engineering* 56, pp. 831-841, 2016.

[46] Chang, Victor, and Gary Wills. "A model to compare cloud and non-cloud storage of Big Data." *Future Generation Computer Systems* 57, pp.56-76, 2016.

[47] Zhang, Xuyun,"A hybrid approach for scalable sub-tree anonymization over big data using MapReduce on the cloud." *Journal of Computer and System Sciences*, pp.1008-1020, 2016.

[48] Assunção, Marcos D,"Big Data computing and clouds: Trends and future directions,**"** *Journal of Parallel and Distributed Computing* 79, pp.3-15, 2015.

[49] Goli-Malekabadi, Zohreh, Morteza Sargolzaei-Javan, and Mohammad Kazem Akbari, "An effective model for store and retrieve big health data in cloud computing**"**, *Computer methods and programs in biomedicine* 132 pp.75-82, 2016.

[50] Malhotra, Rahul, and Prince Jain, "Study and comparison of various cloud simulators available in the cloud computing." *International Journal*, Vol 3, No. 9, 2013.

[51] Xiong, Jinbo, "A secure data self-destructing scheme in cloud computing**”**, *IEEE Transactions on Cloud Computing*, Vol 2, No. 4, pp. 448-458, 2014.

[52] Horváth, Máté, "Attribute-based encryption optimized for cloud computing**”**, *International Conference on Current Trends in Theory and Practice of Informatics*. Springer Berlin Heidelberg, 2015.

[53] Chang, Victor, and Muthu Ramachandran, **“**Towards achieving data security with the cloud computing adoption framework**”**, *IEEE Transactions on Services Computing*, pp.138-151, 2015.

[54] Han, Jinguang,"Privacy-preserving decentralized key-policy attribute-based encryption**”**, *IEEE Transactions on Parallel and Distributed Systems*, pp.2150-2162, 2012.

[55] Li, Ming, "Scalable and secure sharing of personal health records in cloud computing using attribute-based encryption." *IEEE transactions on parallel and distributed systems*, pp.131-143, 2013.

PUBLICATIONS

[1] Jain, B., Bhatia, T., and Verma, A.K. (2017) **"**Security of Big Data in Cloud: A Review**"**, In proceedings of International Conference on Communication, Computing and Networking (ICCCN 2017) pp. 144-150,2017.

[2] Jain, B., Bhatia, T., and Verma, A.K. (2017) **"**ABE based Secure Distributed Storage Scheme for Big Data in Cloud**"**, In proceedings of First International Conference on Smart Technologies in Computer and Communication (SmartTech-2017).

YOUTUBE VIDEO LINK

- https://youtu.be/cu9kNgGcnOh

bhawna_thesis.docx

ORIGINALITY REPORT

10%	8%	6%	6%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

	acme.able.cs.cmu.edu Internet Source	2%
	Submitted Online to Colorado Technical University Student Paper	1%
3	Shimbre, Nivedita, and Priya Deshpande. "Enhancing Distributed Data Storage Security for Cloud Computing Using TPA and AES Algorithm", 2015 International Conference on Computing Communication Control and Automation, 2015. Publication	1%
4	ijcsse.org Internet Source	1%
5	www.qubole.com Internet Source	1%
6	Moghaddam, Faraz Fatemi, Mohammad Ahmadi, Samira Sarvari, Mohammad Eslami, and Ali Golkar. "Cloud computing challenges and opportunities: A survey", 2015 1st	1%

International Conference on Telematics and Future Generation Networks (TAFGEN), 2015.
Publication

7	Submitted to B.S. Abdur Rahman University Student Paper	1%
8	www.sciencedirect.com Internet Source	<1%
9	Submitted to Pondicherry University Student Paper	<1%
10	Li, Yibin, Keke Gai, Longfei Qiu, Meikang Qiu, and Hui Zhao. "Intelligent cryptography approach for secure distributed big data storage in cloud computing", Information Sciences, 2016. Publication	<1%
11	Submitted to Swinburne University of Technology Student Paper	<1%
12	Submitted to Vel Tech University Student Paper	<1%
13	Submitted to The University of the South Pacific Student Paper	<1%
1	Vennila, V., and A. Rajiv Kannan. "Symmetric Matrix-based Predictive Classifier for Big Data computation and information sharing in Cloud",	<1%

Computers & Electrical Engineering, 2016.

Publication