

# International Journal of Advance Research in Computer Science and Management Studies

Research Article / Survey Paper / Case Study

Available online at: [www.ijarcsms.com](http://www.ijarcsms.com)

## *A Smarter Way of Securing and Managing Data for Cloud Storage Applications Using High Throughput Compression in the Cloud Environment*

**Muneshwara M.S<sup>1</sup>**

Assistant Professor  
Department of CS&E

BMS Institute of Technology & Management  
Avalahalli, Yelahanka, Bangalore -560064  
Karnataka – India

**Swetha M.S<sup>2</sup>**

Assistant Professor  
Department of IS&E

BMS Institute of Technology & Management  
Avalahalli, Yelahanka, Bangalore -560064  
Karnataka – India

**Dr. Anil G.N<sup>3</sup>**

Associate Professor  
Department of CS&E

BMS Institute of Technology & Management  
Avalahalli, Yelahanka, Bangalore -560064  
Karnataka – India

*Abstract: Cloud computing has continued to evolve and advance over the ensuing years. Cloud computing is the practice of using a network of remote servers hosted on the internet to manage, store and process data, rather than a local server or a personal computer[1]. This model allows access to information and computer resources from wherever that a network connection is available. Shared pool of resources is being provided which also includes data storage space, computer processing power, networks and specialized corporate and user applications. By moving to the cloud many companies achieved operational savings, with 88 percent of companies surveyed saying the cloud has reduced costs. In addition, data storage or data retrieval cost is high especially for small companies[2]. Economic computing resources and advanced network technology is referred to as cloud computing .The use of cloud computing has increased rapidly in many organizations.*

*Security is not guaranteed since the data placed in the cloud is accessible to everyone. Cryptographic techniques cannot be directly adopted to ensure security. To maintain the reputation, the cloud service provider may hide the data corruptions sometimes. For this we present a new approach that facilitates to reduce the cost than normal usage and providing much security also. In order for data to be effectively managed we follow the most well-known area in the IT, “Data Compression “which is the process of reducing the size of a data file [2]. We can expect Bandwidth to either be limited, expensive, or both, and data compression helps to alleviate these bottlenecks. Data compression squeezes data so it requires less disk space for storage and less bandwidth on a data transmission channel. By using this technique we can reduce the size of data stored in the cloud and along with this we can provide sufficient security by providing another level of security by proxy re-encryption and key policy attribute based encryption.*

*The survey demonstrates that cost is a huge factor for companies that want to realize high returns on their IT investments. To save even more money through the cloud for companies is to compress their data before it goes into storage. The size of large files can be reduced so that they take up a fraction of the space using data compression [2][3]. Because many companies pay according to the amount of capacity used, reducing the size of files before making a cloud migration can result in large savings.*

**Keywords:** Bandwidth; Cloud computing; Cryptography; Data Compression; Decryption; Encryption.

## I. INTRODUCTION

As learned from past events, computing in its purest form has changed hands multiple times. Mainframes were predicted to be the future of computing in the near beginning. Indeed mainframes and large scale machines were built and used in those days and in some circumstances are used similarly today [2]. The bigger and more expensive turned to smaller and more affordable commodity PCs and servers where there was a change in trend.

Most of our data is stored on local networks with servers that may be clustered and sharing storage. This approach developed into stable architecture and provides decent redundancy when it is deployed right. A demanding attention is drawn by a newer emerging technology, cloud computing and quickly is changing the direction of the technology landscape. Cloud computing is a playing field leveler; it gives small businesses access to technologies that previously were out of their reach and lets small businesses compete with both other small businesses and larger ones[3].

Technology is a funny thing, just when you think that it has reached its highest point it will surprise you by showing something more amazing than you thought possible. Technology has thrown our way to win us over and cloud computing is one of such trump cards. Truly a thing of past because of cloud computing is the times when corporations and organizations bought new software or license for new workers [4]. A nicely organized user interface lets multiple users access same software. Work flow is made smoother by cutting down the expenses of big companies. The remote servers you hire for yourself or your business accommodates your need for superior bandwidth anytime you need it. Professionals finish their work faster and finish the work within the stipulated time with the help of cloud computing. We often fear that our hard work and saved files will get destroyed somehow and take elaborate precautions. The worries of safeguarding our important files and data are taken away by cloud computing.

### A. Overview

The rapid adoption of cloud based applications has increased a demand for a robust cloud data management solution providing the platform and infrastructure that has resulted in more fragments of data to be scattered both inside and outside of the firewall across the enterprise.

As promising as it is, cloud computing is also facing many challenges that, may impede its fast growth, if not well resolved [1]. The most complicated topic today is the issues surrounding the cloud. The reality is that the lifeblood of organizations is the data. Therefore, regardless of where actually the data lives, managing it is very important [5]. Data is created, changed, secured, stored and governed which defines the entire life cycle of data management.

Cloud invisibly backs up the files and folders and elevates the potentially endless and costly search for extra storage space from music files to pictures to sensitive documents. Using cloud storage services means that you and others can access and share files across a range of devices and locations. Files such as photos and videos can sometimes be difficult to email if they are too large or you have a lot of them. We can quickly circulate a URL and can share the files with anyone we choose, which would mean uploading to the cloud storage.

Management of data becomes paramount in maintaining service levels and securing the critical business information because all data in a cloud lives in the same shared system [6]. Cloud computing enables to be excessively dependent on the internet. The availability of the robust and reliable internet for all the time is the premise on which the cloud computing exists.

For the purpose of upholding the efficiency and effectiveness of cloud computing in data management even in the case of reduced bandwidth and irrespective of the device used to retrieve data, a feasible solution could be to compress the data on cloud.

Data compression helps to reduce the consumption of expensive resources in this regard. For example, the disk space or transmission bandwidth.

## II. LITERATURE SURVEY

The following describes about all the preliminary concepts and prior work done in the similar area.

### A. Cryptography

Cryptography is the science of writing in secret code using mathematics by encompassing the principles and methods of transforming an intelligible message into one that is unintelligible (encrypting), and then retransforming that message back to its original form (decrypting). In data and telecommunications, cryptography is necessary when communicating over any untrusted medium, which includes just about any network, particularly the Internet [1]. As the Internet and other forms of electronic communication become more prevalent, electronic security is becoming increasingly important.

There are some specific security requirements, including:

**Authentication:** The process of proving one's identity. (The primary forms of host-to-host authentication on the Internet today are name-based or address-based, both of which are notoriously weak).

**Privacy/confidentiality:** Ensuring that no one can read the message except the intended receiver.

**Integrity:** Assuring the receiver that the received message has not been altered in any way from the original.

**Non-repudiation:** A mechanism to prove that the sender really sent this message.

There are, in general, three types of cryptographic schemes typically used to accomplish these goals:

1. Secret Key Cryptography (SKC): Uses a single key for both encryption and decryption
2. Public Key Cryptography (PKC): Uses one key for encryption and another for decryption
3. Hash Functions: Uses a mathematical transformation to irreversibly "encrypt" information.

### B. Symmetric key Cryptography

In secret key cryptography, a single key is used for both encryption and decryption. The sender uses the key to encrypt the plaintext and sends the cipher text to the receiver. The receiver applies the same key to decrypt the message and recover the plaintext.

Because a single key is used for both functions, secret key cryptography is also called symmetric encryption. With this form of cryptography, it is obvious that the key must be known to both the sender and the receiver; that, in fact, is the secret. The biggest difficulty with this approach, of course, is the distribution of the key. Secret key cryptography schemes are generally categorized as being either stream ciphers or block ciphers. Secret key cryptography algorithms that are in use today include: Data Encryption Standard and Advanced Encryption Standard.

### C. Public key cryptography

Public or asymmetric key cryptography involves the use of key pairs: one private key and one public key. Both are required to encrypt and decrypt a message or transmission. With public key cryptography, all parties interested in secure communications publish their public keys. The private key, not to be confused with the key utilized in private key cryptography, is just that, private. It is not to be shared with anyone

### D. Lossless Compression

All of the information is completely restored with lossless compression, where every single bit of data that was originally in the file remains after the file is decompressed. It is used for files, such as applications that need the original file. This is generally the technique of choice for spreadsheet files or text, where losing words or financial data could pose a problem [7]. The basic principle that lossless compression algorithms work on is that any non-random file will contain duplicated

information that can be condensed using statistical modeling techniques that determine the probability of a character or phrase appearing. These statistical models can then be used to generate codes for specific characters or phrases based on their probability of occurring, and assigning the shortest codes to the most common data.

Lossless Compression techniques include Run-length encoding, entropy encoding, Huffman coding, Lempel-Ziv algorithms [5] [7].

### III. PROBLEM DEFINITION

The most complicated topic today is the issues surrounding the cloud. Management of data becomes paramount in maintaining service levels and securing the critical business information because all data in the cloud lives in the same shared system. The reality is that the lifeblood of the organization is the data. Data compression helps to reduce the consumption of expensive resources in this regard. Therefore, regardless of where actually the data lives, managing it is very important.

### IV. SYSTEM ANALYSIS

Systems analysis is the study of sets of interacting entities. System analysis is "the process of studying a procedure or business in order to identify its goals and purposes and create systems and procedures that will achieve them in an efficient way".

#### A. Existing System

Cloud computing promises to increase the velocity with which applications are deployed, all while increasing business agility, increase innovation, and lower costs. The architecture of cloud computing can be divided into two sections: the front end and the back end. They connect to each other through the network, which would usually be the Internet.

**Front end:** The front end includes the client's computer (or computer network) and the application required to access the cloud computing system. Some user interface is not provided in all cloud computing systems [6][9]. Services like Web-based e-mail programs leverage existing Web browsers for example like Internet Explorer or Firefox. Other systems have unique applications that provide network access to clients.

**Back end:** There are various computers on the back end of the system, servers and data storage systems that create the "cloud" of computing services. A cloud computing system could include practically any computer program we can imagine, from data processing to video games in theory. Usually, each application will have its own dedicated server.

A central server administers the system, monitoring traffic and client demands to ensure everything runs smoothly. It follows a set of rules called protocols and uses a special kind of software called middleware.

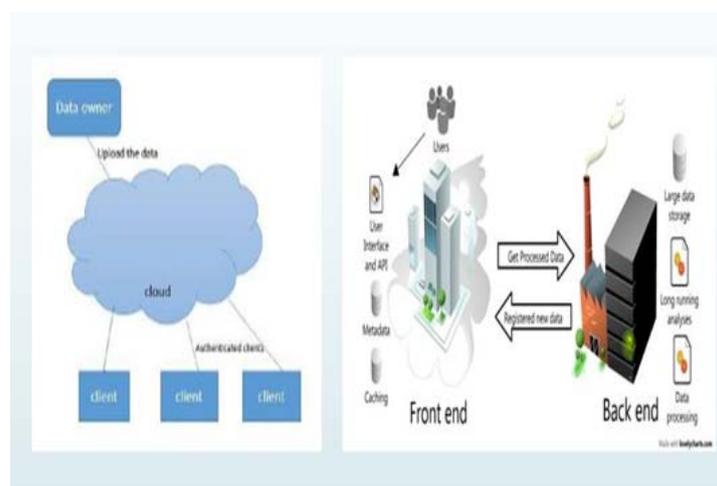


Figure 1: The Existing System

The working of cloud can be illustrated as follows:

There are an increasing number of services offering 'cloud storage' where you can upload documents, photos, videos and other files to a website to share with others or to act as a backup copy. These files can then be accessed from any location or any

type of device (laptop, mobile phone, tablet etc.). Once you have registered for an account you typically create a folder on your computer and every file you place in that folder is copied to the servers of the storage provider [9]. Any changes made to these files are automatically copied across and immediately accessible from other devices you may have. If you choose to store your files in the cloud you need to remember that this means they are really just stored on servers controlled by the service provider. Some providers of cloud services may also use the cloud services of another organization. This means you need to check that the security and availability of the service is right for the types of files you want to upload. Using cloud storage services means that you and others can access and share files across a range of devices and locations. Files such as photos and videos can sometimes be difficult to email if they are too large or you have a lot of them. Uploading to a cloud storage provider means you can quickly circulate a URL and you can share your files with anyone you choose.

## B. Proposed System

There are many challenges in Cloud Computing, if not well resolved, may impede its fast growth. In this work the challenges of the existing techniques are addressed by using compression technique and the security issues by some techniques of Key-policy attribute-based encryption (KP-ABE) and Proxy re-encryption. Compression technique is proposed to manage data efficiently on cloud to reduce the storage space. KP-ABE is proposed to resolve the problem of data access control in one-to-many communications, while proxy re-encryption is a technique which is mainly used to protect the data from its disclosure in the cloud storage server. This technique helps the cloud servers to provide the requested data to the client without knowing the underlying content.

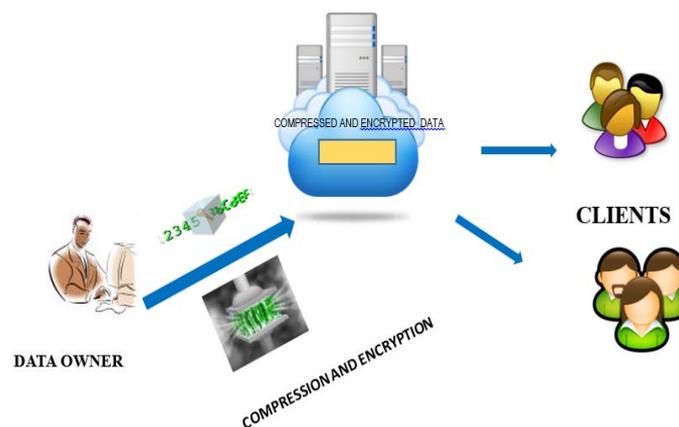


Figure 2: The Proposed System

The major entities are:

**Data owner:** An entity, who has large data files to be stored in the cloud and relies on the cloud for data maintenance and computation, can be either individual consumers or organizations.

**Cloud Storage Server (CSS):** An entity, which is managed by Cloud Service Provider (CSP). Client's data is maintained by the significant storage space and computation resource.

**Clients:** Authenticated clients who can access information from the cloud in a secure way.

Data compression involves encoding information using fewer bits than the original representation. Compression can be of two types: lossy or lossless. Lossless compression decreases the number of bits by identifying and eliminating statistical redundancy. Information is not lost in lossless compression. Lossy compression will reduce bits by identifying unnecessary information and removing it. Although the formal name is source coding, the process of reducing the size of a data file is popularly referred to as data compression. Multimedia files are large and consume lots of hard disk space. In order to distribute over the internet, the files size makes it time-consuming to move them from place to place over school networks. Compression shrinks files, making them smaller and more practical to store and share. Compression works by removing repetitious or redundant information, effectively summarizing the contents of a file in a way that preserves as much of the original meaning as

possible. At any given time, the ability of the Internet to transfer data is fixed with a view of the capability as the Internet's collective bandwidth.

Thus, if data can effectively be compressed wherever possible, significant improvements of data throughput can be achieved. In some instances, file sizes can be reduced by up to 60-70 %. At the same time, many systems cannot accommodate purely binary data, so encoding schemes are also employed which reduce data compression effectiveness. The compression technique used in this work is a combination of LZ77 and Huffman Coding.

## V. SYSTEM DESIGN

The data owner is the one who compresses and encrypts the document (text, image and word document) that he wants to upload on the cloud server. After which he performs the AES encryption to encrypt the plain document, proxy re-encryption and attribute based encryption to provide unique access structure to each client. Then he categorizes the data into different categories by linking different proxy key generated using the proxy re-encryption technique[10]. The data owner is the one who has the priority of editing and also deleting the contents stored on the cloud server. Then he performs the attribute based encryption to provide unique access structure to each client. After all these steps the data owner uploads the compressed and encrypted file, re-encryption key and access structure on to the cloud server. Only authorized client can log in using his private key. After successful he gets a unique access structure that has been assigned by the data owner. Clients request the cloud server for the re-encryption key and files that he wants to download. Finally client decrypts the re-encryption key and then decrypts the compressed and encrypted file using this key.

The system architecture is as follows:

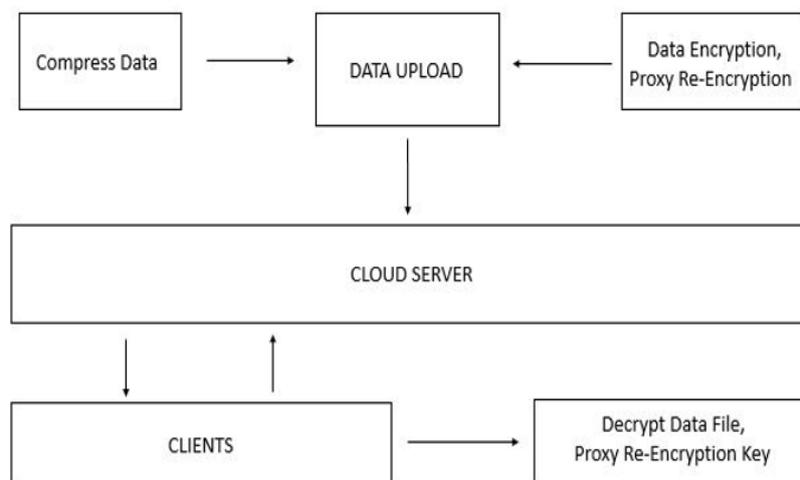


Figure 3: The block diagram of the proposed system.

The data owner is the one who compresses and encrypts the document (text, image and word document) that he wants to upload on the cloud server. After which he performs the AES encryption to encrypt the plain document, proxy re-encryption and attribute based encryption to provide unique access structure to each client. Then he categorizes the data into different categories by linking different proxy key generated using the proxy re-encryption technique. The data owner is the one who has the priority of editing and also deleting the contents stored on the cloud server. Then he performs the attribute based encryption to provide unique access structure to each client. After all these steps the data owner uploads the compressed and encrypted file, re-encryption key and access structure on to the cloud server. Only authorized client can log in using his private key. After successful he gets a unique access structure that has been assigned by the data owner. Clients request the cloud server for the re-encryption key and files that he wants to download. Finally client decrypts the re-encryption key and then decrypts the compressed and encrypted file using this key.

## A. Detailed Design

A sequence diagram is an interaction diagram that shows how processes operate with one another and in what order. A construct of a Message Sequence Chart is the design in detail. A sequence diagram shows object interactions arranged in time sequence in an orderly manner. It illustrates the objects and classes involved in the scenario and the sequence of messages exchanged between the objects needed to carry out the functionality of the scenario.

A sequence diagram is a UML diagram that provides a graphical view of the chronology of the exchange of messages between objects and actors for a use case, the execution of an operation, or an interaction between classes, with an emphasis on their chronology. Sequence Diagrams are used primarily to design, document and validate the architecture, interfaces and logic of the system by describing the sequence of actions that need to be performed to complete a task or scenario.

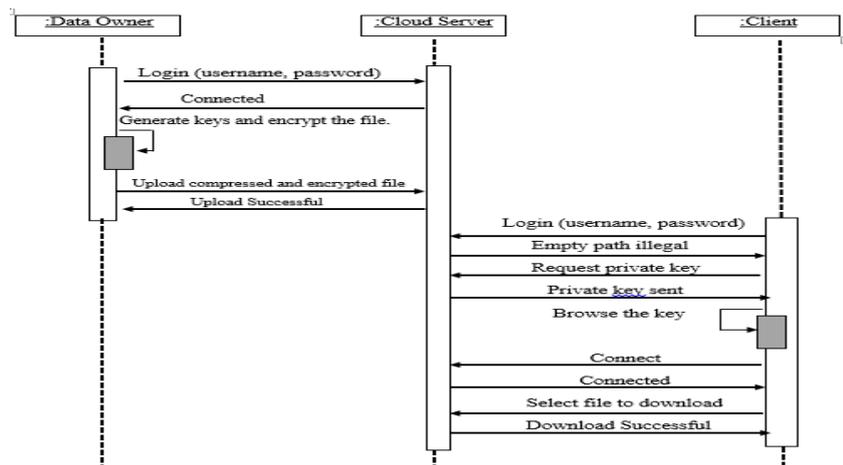


Figure 4: The sequence diagram

## VI. SYSTEM IMPLEMENTATION AND RESULTS

Implementation is a state where theoretical design turns into working system. Achieving a new successful system and giving confidence in new system that it will work effectively and efficiently is the most dynamic stage. The system is implemented only after thorough checking is done and if it is found working in according to the specifications. Implementation of any software or system is always preceded by important decisions regarding selection of the platform used to implement, the language used to code, etc. These decisions are often influenced by several factors such as the real environment in which the system works[11], the speed that is required, the security concerns, other implementation specific details etc.,

### A. Algorithm for the overall system is as follows:

**Input:** Text, Word or Pdf document, video, audio and images.

**Output:** Text, Word or Pdf document, video, audio and images.

**Step 1:** Initially the cloud server is started.

**Step 2:** Then the data owner login using the data upload interface.

**Step 3:** Data owner compresses the data.

**Step 4:** He encrypts the data.

**Step 5:** Performs PRE (Proxy Re-Encryption), using client's public key.

**Step 6:** Further he performs KP-ABE (Key Policy Attribute Based Encryption).

**Step 7:** He then generates different categories using the proxy key generated during the PRE, Upload the data.

**Step 8:** Authorized clients who are authenticated by the data owner login using the requested private key.

**Step 9:** Download required files.

The data owner is upload the data or document (text, image and word document) to the cloud server and client will access the data by using is personal user name and the passwords.

Browse the compressed and encrypted file to upload and click upload



Figure 5: Upload File to Server

If user is given access structure in the server automatically list of category will be displayed on the left side of the panel. Download link is available on the last column of each records and click on it to decrypt the file using proxy re-encryption key as shown in following screenshot, shows the completion of download.

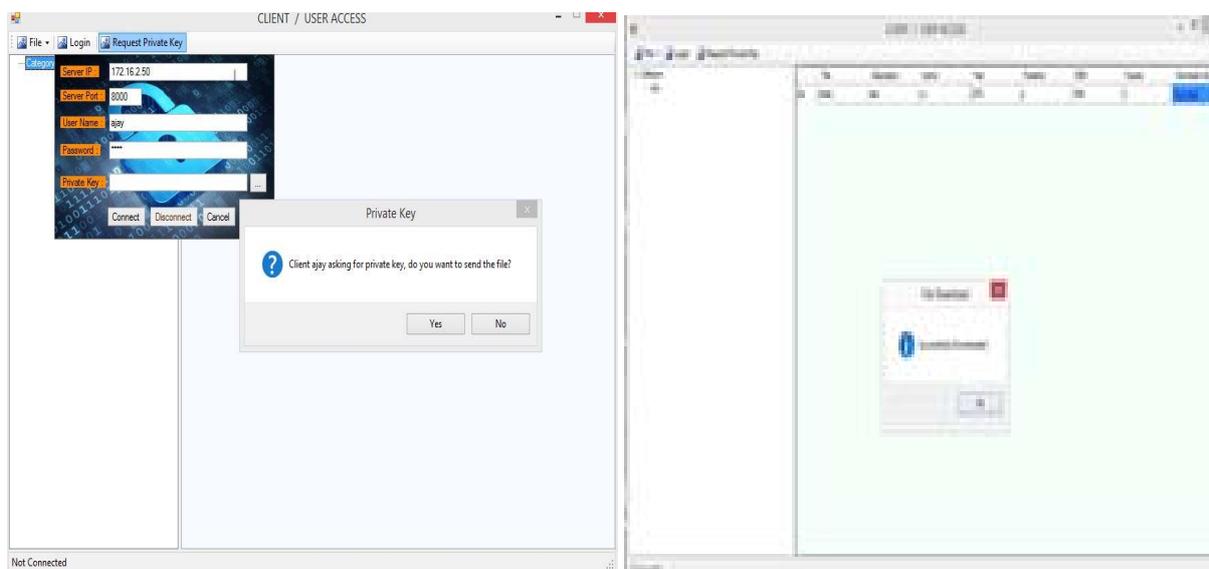


Figure 6: Client Downloaded File

## VII. CONCLUSION AND FUTURE WORK

This paper work studies the problem of ensuring the integrity of data storage in Cloud Computing. We outline the challenges associated with the retrieval of data from cloud in an appropriate manner. As the data gets compressed, it leads to a more optimized way of retrieving data from cloud. The use of compression in cloud computing leads to effective use of storage disks and bandwidth. This work enables the user to fine-tune the trade-off between storage costs, computation time and bandwidth costs. Different computations of characters can be represented by fewer number of bits in compression, which is an efficient way of retrieving data in the cloud environment.

The technique of compression, AES, RSA and KP-ABE algorithm are used to monitor the tasks simultaneously. Unlike most prior works, the new scheme further supports secure and efficient dynamic operations on data stored in the cloud,

including: data compression, data encryption, data update, delete and append. Unique access structure enhances security. Implementation simplifies the development efforts and provides direct support for security where data can be managed effectively in cloud. Extensive security and performance analysis shows that the proposed scheme is highly efficient and provably secure.

In future, this work can be enhanced by developing a module where data can be retrieved in an efficient way based on the Internet bandwidth. An attempt to reduce the compression speed can be done to improve the effectiveness of future implementation activities.

#### ACKNOWLEDGEMENT

The authors would like to thank the editor, mystifying reviewers for their priceless suggestions that appreciably improved the eminence of this paper.

I heartily thank Principal, B M S Institute of Technology & Management, Bangalore, HOD, Department of CSE, B M S Institute of Technology & Management, Bangalore. Finally thankful to our dear faculties, Department of CSE B M S Institute of Technology & Management, Bangalore, affiliated to Visvesvaraya Technological University, Belgaum, KARNATAKA, INDIA.

I Special thank Dr. Anil G.N for their continuous encouragement and motivation. Last but not least to all of them who helped us to get a warm success in completion of this paper.

#### References

1. Bicer, T.; Comput. Sci. & Eng., Ohio State Univ., Columbus, OH, USA ; Jian Yin ; Chiu, D. ; Agrawal, G. ” **Integrating Online Compression to Accelerate Large-Scale Data Analytics Applications**”, Parallel & Distributed Processing (IPDPS), 2013 IEEE 27th International Symposium.
2. Abdelaal, Mohamed ; System Software and Distributed Systems, Carl von Ossietzky University of Oldenburg, Germany : “**An efficient and adaptive data compression technique for energy conservation in wireless sensor networks**”, Wireless Sensor (ICWISE), 2012 IEEE Conference.
3. Samlinson, E. ; Department of Computer Science and Engineering, Sona College of Technology, Salem, India ; Usha, M. : “**User-centric trust based identity as a service for federated cloud environment**”, Computing, Communications and Networking Technologies (ICCCNT),2013 Fourth International Conference.
4. Bandeira Soares, L. ; Inf. Inst. - Microelectron. Grad. Program (PGMicro), Fed. Univ. of Rio Grande do Sul (UFRGS), Porto Alegre, Brazil ; Bampi, S. ; David, C. ;Crovato, P. : “**A fast EMD-based technique for Power Quality signals decomposition, compression, and time-frequency analysis**”, Digital Signal Processing (DSP), 2010 18th International Conference.
5. Lorca, J. ; Transversal Projects & Innovation, Telefonica I+D, Madrid, Spain ; Cucala, L. : “**Lossless compression technique for the fronthaul of LTE/LTE-advanced cloud-RAN architectures**”, World of Wireless, Mobile and Multimedia Networks (WoWMoM), 2013 IEEE 14th International Symposium and Workshops.
6. Qian Wang, Cong Wang, Jin Li, Kui Ren, and Wenjing Lou,” **Enabling Public Audit ability and Data Dynamics for Storage Security in Cloud Computing**”, IEEE computer society , Vol 22, No 5, May 2011
7. M. Armbrust, A. Fox, R. Griffith, A. D. Joseph, R. H. Katz, A. Konwinski, G. Lee, D. A. Patterson, A. Rabkin, I. Stoica, and M. Zaharia, “Above the clouds: **A Berkeley view of cloud computing**”, University of California, Berkeley, Tech. Rep. USB-EECS-2009-28, Feb 2009.
8. Qian Wang, Cong Wang, Jin Li, Kui Ren, and Wenjing Lou, “**Enabling Public Verifiability and Data Dynamics for Storage Security in Cloud Computing**”, Springer-Verlag Berlin Heidelberg 2009.
9. Vipul Goyal”**Attribute-Based Encryption for Fine-Grained Access Control of Encrypted Data**”CCS’06, October 30–November 3, 2006,
10. Ting Yu” **A Unified Scheme for Resource Protection in Automated Trust Negotiation**” IEEE 2003.
11. Carpen-Amarie, A.INRIA/IRISA, Rennes, Towards a **Self-Adaptive Data Management System for Cloud Environments**.
12. Istanbul Tech. Univ., Istanbul, **Multi Stage Vector Quantization for the Compression of Surface and Volumetric Point Cloud Data**.
13. Yijin Chen ; Coll. of Geosci. & Surveying Eng., China Univ. of Min. & Technol., Beijing, China ; Huixia Zhang ; Xiaoxue Fu, Organization and query of point clouds data based on SQL Server spatial.
14. Meng-Ju Hsieh ; Chao-Rui Chang ; Li-Yung Ho ; Jan-Jan Wu ;Pangfeng Liu, SQLMR :**A Scalable Database Management System for Cloud Computing**.
15. Cho Cho Khaing ; Thinn Thu Naing, **The efficient data storage management system on cluster-based private cloud data center**.
16. Anne, V.P.K. ; Ponnam, V.S. ; Praveen, G. Software Engineering (CONSEG), 2012 CSI Sixth International Conference on **A significant approach for cloud database using shared-disk architecture**.
17. Ooi Beng Chin ; Nat. Univ. of Singapore, **Cloud Data Management Systems:Opportunities and Challenges**.

**AUTHOR(S) PROFILE**

**Mr. Muneshwara M.S** is an Assistant Professor in the Department of Computer Science and Engineering at B M S Institute of Technology & Management, Bangalore, affiliated to Visvesvaraya Technological University, KARNATAKA, INDIA He has Completed M. Tech in Computer Science & Engineering Branch & B.E. Degree in Information Science & Engineering Branch from Visvesvaraya Technological University, Belgaum, KARNATAKA,INDIA. He has around 9 years of experience in teaching.



**Mrs. Swetha M.S** is an Assistant Professor in the Department of Information Science and Engineering at B M S Institute of Technology & Management, Bangalore, affiliated to Visvesvaraya Technological University, KARNATAKA, INDIA. She has Completed M. Tech in Computer Science & Engineering Branch Visvesvaraya Technological University, KARNATAKA. & B.E. Degree in Computer Science & Engineering Branch from Visvesvaraya Technological University, KARNATAKA and She has around 5 years of experience in teaching.



**Dr. Anil G.N** is an Associate Professor in the Department of Computer Science and Engineering at B M S Institute of Technology & Management, Bangalore, affiliated to Visvesvaraya Technological University, KARNATAKA, INDIA. He has completed Doctor of Science (D.Sc) from Rani Channama University, Belgaum. Karnataka, India, M. Tech in Computer Science & Engineering Branch from OUCE, Osmania University, Hyderabad, India & B.E. Degree in Computer Science & Engineering Branch from Mysore University and He has around 19 years of experience in teaching & 7 years of experience in research.