Cluster Based Multi Layer User Authentication Data Center Storage Architecture for Big Data Security in Cloud Computing

S. Ramasamy[1], R. K. Gnanamurthy[2]

[1] Department of CSE, Vivekanandha College of Technology for Women, India
[2] Department of ECE, Dhanalakshmi Srinivasan College of Engineering, India
ramasamycse85@gmail.com, rkgnanam@yahoo.co.in

Abstract

In distributed computing, cloud is a fast emergent technique in data analytic technology era which is applied to stock up and retrieves the big data in distributed environment. Every day an individual person and companies are placing more data in the cloud. The organization authorities and individual users are started to worry about how the big data are safe in cloud. Information protection is one of the foremost troubles, which reduce the development of cloud computing and make difficult by means of data confidentiality and security. Day by day researchers has been introducing new techniques for secure big data in cloud computing environment. In conventional security systems are used encryption and decryption for providing security in cloud data. In our proposed work we use multilayer user authentication mechanism for big data security. First, we deal about the safety measures and challenges in cloud computing. Second, we propose cluster based multi layer user authentication data centre storage design for defending big data in cloud computing atmosphere. Multilayer authentication provides information security in three layers with different defence factors. Service providers categorize the data units in to three different clusters based on the usage of data. Proposed systems ensure the big data security through authentication in cloud and get extra engineering insights.

Keywords: Cloud computing, Security, Big data, Cluster, Multilayer

1 Introduction

1.1 Cloud Computing

According to U.S National Institute of Standards and Technology (NIST), Cloud Computing is a method for accessing favourable, on demand services to a collective pool of customisable computing properties that can be easily adoptable and unconfined with negligible organization attempt or cloud supplier communication [1].

Cloud computing is an emerging new computing paradigm for delivering computing services. This computing approach relies on a number of existing technologies, e.g., the Internet, virtualization, grid computing, Web services, etc. Cloud Computing aims to provide scalable and inexpensive on-demand computing infrastructures with good quality of service levels. It represents a shift away from computing as a product that is purchased, to computing as a service that is delivered to consumers from the cloud. It helps an organization in saving costs and creating new business opportunities [24].

Cloud computing has formed the conceptual and infrastructural basis for tomorrow's computing. The global computing infrastructure is rapidly moving towards cloud based architecture. While it is important to take advantages of could based computing by means of deploying it in diversified sectors, the security aspects in a cloud based computing environment remains at the core of interest [25].

Cloud based services and service providers are being evolved which has resulted in a new business trend based on cloud technology. With the introduction of numerous clouds based services and geographically dispersed cloud service providers, sensitive information of different entities are normally stored in remote servers and locations with the possibilities of being exposed to unwanted parties in situations where the cloud servers storing those information are compromised. If security is not robust and consistent, the flexibility and advantages that cloud computing has to offer will have little credibility [25].

Cloud Computing is the highly emerging area in the computer world now a days by which the web applications are delivering the computing services through internet on demand. Cloud computing making business applications more mobile, which is of a great concern and making the cloud computing more popular and necessity for the large business organization to grow more rapidly [26].

*Corresponding Author: S. Ramasamy; E-Mail: ramasamycse85@gmail.com
DOI: 10.3966/160792642020012101015

1.1.1 Characteristics of Cloud Computing

The following Figure 1 shows the various characteristics of cloud computing.

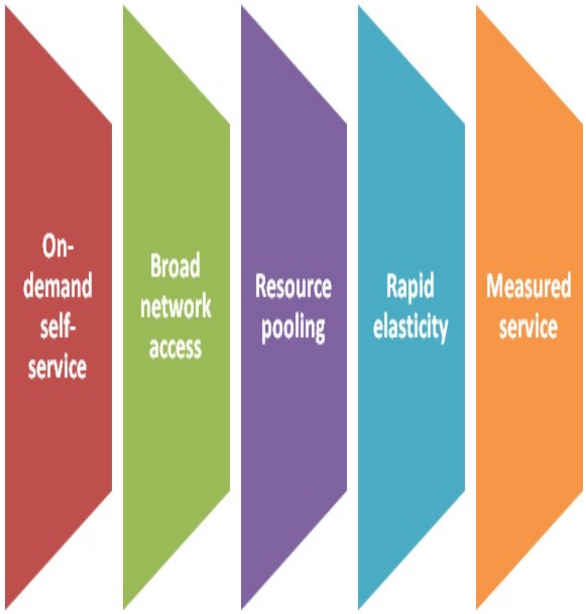

Figure 1. Cloud computing characteristics

On demand self service. A customer can unilaterally stipulation computing capability, such as system time, set of connections and storage, as required mechanically not including requiring individual communication with each service contributor [1].

Industrialized association can stipulation supplementary computing possessions as essential lacking available all the way through the cloud facility supplier, this can be a storeroom space. Modern organizations can make use of a web portal as an boundary to access their cloud financial records to see their cloud services, their usage, and also to stipulation and de-prerequisite services as they require to.

Broad network access. Capabilities are accessible over the set-up and accessed all the way through standard mechanisms that support use by mixed skeletal or broad client platforms e.g., cellular phone, tablets, notebook, and desktop [1].

System bandwidth and latency are significant aspects of cloud and broad set of connection access, because they communicate to the quality of service (QoS) on the set of connections. This is a main key for allocating moment in time sensitive industrialized applications.

Multi tenancy and Resource pooling. Cloud computing properties are intended to sustain a multi-tenant model. Multi-tenancy allows many consumers to divide up the same applications or the same substantial infrastructure while retaining isolation and protection over their data's.

The suppliers computing possessions are pooled to serve various consumers by means of a multi-tenant representation, with dissimilar physical and essential resources with dynamism assigned and reassigned as reported by consumer demand. There is an intelligence of location autonomy in that the consumer generally has no rule or familiarity over the exact position of the provided possessions but may be able to identify location at a superior level of generalization (e.g., city, state, or data unit). Examples of possessions include memory, processing, system and network capacity [1].

Resource pooling means that multiple customers are serviced from the same physical resources. Service provider resource pools have to be extremely bulky and elastic sufficient to service many client needs and to make available for financial system of scale. While it comes to resource pooling, resource allotment have to not impact performances of significant developed applications.

Rapid elasticity and Scalability. Capacity can be easily provisioned and on the loose, in some cases routinely, to scale quickly external and inward corresponding with demand. To the customer, the capacities available for provisioning frequently appear to be unconstrained and can be appropriated in any amount at every time [1].

Elasticity is a familiar sight of cloud computing and it implies that mechanized organizations can rapidly provision and de-provision any of the cloud computing resources. Rapid provisioning and de-provisioning might apply to storage or virtual machines or customer applications.

Cloud computing scalability, there is a smaller amount assets expenses on the cloud consumer side. This is since as the cloud consumer needs extra computing resources, they can simply provision them as needed, and they are available right away. Scalability is way that developed organizations are progressively planning for more capability and of itinerary the cloud can hold that scaling high or scaling down.

Measured service. Cloud systems mechanically organize and optimize source use by leveraging a measured capability at various level of concept suitable to the type of facility (e.g., memory, dispensation, network capacity, and active user financial records). Source usage can be guarded, monitored, and reported, provided that intelligibility for both the supplier and customer of the utilized facility [1].

Cloud computing property usage is metered and modern organizations reimburse consequently for what they have use. Source consumption can be optimized by leveraging pay-per-use capability. This is the way that cloud source usage in virtual server instances that are successively gets monitored, calculated and reported by the cloud service provider.

1.1.2 Services of Cloud Computing

The following Figure 2 shows the various services offered by the cloud computing.

Figure 2. Services offered by cloud computing

Software as a Service (SaaS). The capacity offered to the customer is to use the supplier's applications organization on cloud transportation. The applications are easily reached from a variety of client strategy through whichever a skeletal customer interface, such as a web browser (e.g., web-based electronic mail), or a program crossing point. The customer does not administer or manage the fundamental cloud transportation as well as communication network, operating systems, memory, servers, or even entity application capabilities, with the possible prohibiting of limited user exact application constitution settings [1].

Platform as a Service (PaaS). The potential provided to the customer is to organize onto the cloud transportation consumer produced or offered applications formed using programme languages, libraries, forces, and tools supported by the supplier. The customer does not deal with or control the fundamental cloud transportation as well as servers, communication network, functioning systems, or memory, but has manage over the deployed applications and probably constitution settings for the application hosting surroundings [1].

Infrastructure as a Service (IaaS). The ability provided to the end user is to prerequisite dispensation, networks, memory and other essential computing possessions where the customer is able to organize and run random software, which can contain functional systems and applications. The customer does not administer or organize the fundamental cloud transportation but has organize over working systems, deployed applications and memory; and probably limited organize of select communication networking components [1].

1.1.3 Types of Cloud Computing

Private cloud. The cloud transportation is provisioned for private use by a particular association comprising numerous clients (e.g., industry units). It may be owned, organised, and operated by the association, a third party, or some grouping of them, and it may perhaps continue living on or off property [1].

Community cloud. The cloud computing is provisioned for special use by an exact centre of population of customers from organizations that contain mutual concerns (e.g., undertaking, safety requirements, course of action, and conformity considerations). It may be owned, organised, and maintained by one or more of the industries in the society, a third party, or some grouping of them, and it may possibly continue living on or off premises [1].

The above Figure 3 describes the different types of cloud computing.

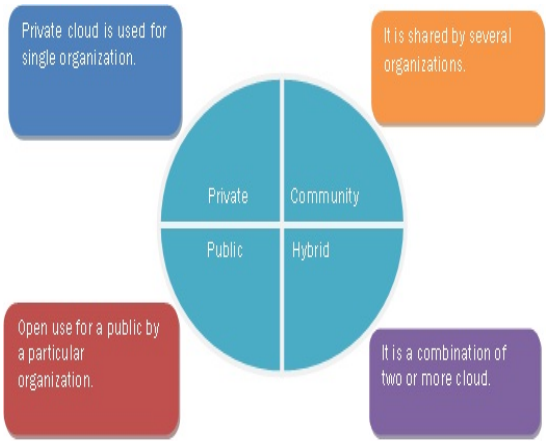

Figure 3. Types of Cloud Computing

Public cloud. The cloud transportation is to supply for open make use of by the common community. It may be organised, owned and operated by a company, educational, or government association, or some grouping of them. It exists on the property of the cloud supplier [1].

Hybrid cloud. The cloud computing transportation is a combination of two or more different cloud organizational facilities (private, public and community) that stay on exclusive entities, but are jump jointly by harmonized or proprietary knowledge that enables information and function compatibility (e.g., cloud convulsive for load evaluation between clouds) [1].

1.2 Big Data

Big data is an expression for information collections that are so enormous that conservative data dispensation request software is insufficient to agreement with them. Big data challenges comprise collecting data, information storage, information analysis, and penetrating, partaking, transformation, hallucination, querying, and updating data privacy.

The phrase huge data disposition to relocate the use of prophetic analytics, user performance analytics, or convinced previous highly residential information analytics models that receive missing value from information, and infrequently to a meticulous size of

records set. Data mining tools predict behaviours and future, predisposition allowing businesses to make proactive, knowledge driven decisions. Big data traditionally includes set of information's with sizes away from the possible of frequently used software equipment to incarcerate, curate, supervise, and development data within a manageable ancient history time [23].

The above Figure 4 describes the different components of big data. The big data term which is being send now a days is kind of misnomer as it points out only the size of the data not putting too much of attention to its other existing properties. Big data can be defined with the following properties associated with it [2]. In Figure 5 we deal with the three different Vs of big data volume, velocity and variety. These three Vs are explained briefly in the following steps.

Figure 4. Big data

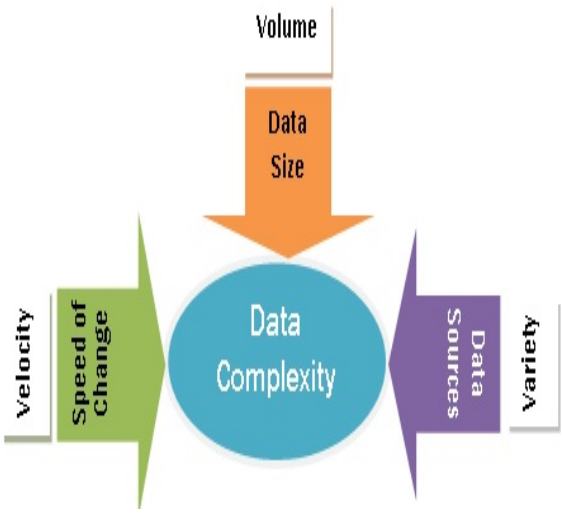

Figure 5. Three V's of big data

Variety. Data being produced is not of single category as it not only includes the traditional data but also the semi structured data from various resources like web Pages, Web Log Files, social media sites, e-mail, documents, sensor devices data both from active passive devices. All this data is totally different consisting of raw, structured, semi structured and even unstructured data which is difficult to be handled by the existing traditional analytic systems [2].

Volume. The Big word in big data itself defines the volume. At present the data existing is in peta bytes and is supposed to increase to zetta bytes in nearby future. The social networking sites existing are themselves producing data in order of terabytes everyday and this amount of data is definitely difficult to be handled using the existing traditional systems [2].

Velocity. Velocity in Big data is a concept which deals with the speed of the data coming from various sources. This characteristic is not being limited to the speed of incoming data but also speed at which the data flows. For example the data from the sensor devices would be constantly moving to the database store and this amount won't be small enough. Thus our traditional systems are not capable enough on performing the analytics on the data which is constantly in motion [2].

The other two dimensions that need to consider with respect to Big Data are Variability and Complexity [2].

Variability. Variability considers the inconsistencies of the data flow. Data loads become challenging to be maintained especially with the increase in usage of the social media which generally causes peak in data loads with certain events occurring [2].

Complexity. It is quite an undertaking to link, match, cleanse and transform data across systems coming from various sources. It is also necessary to connect and correlate relationships, hierarchies and multiple data linkages or data can quickly spiral out of control [2].

1.3 Security Challenges in Cloud Computing

Cloud computing is a collection of services and applications including individual and business. The services like software, platform and infrastructure. Applications like web application, utility application and data applications, these applications and services are offered at entire world.

There are plentiful safety issues for cloud computing as it encounter numerous technology includes networking, data mining and analytics , operational systems, virtualization, source scheduling, operation management, shipment balancing, utility control and storage management. Therefore, safety issues for numerous of these systems and technology is appropriate to cloud computing. For example, a network that interconnecting two or more systems in a cloud comprise to be protected and linking the virtual machinery to the physical machinery has to be approved out strongly.

The above Figure 6 defines the various security issues in the cloud computing. Data security involves authenticating the user to access data in addition as ensuring that proper rules and regulations are imposed

for data distribution. The following are the variety of safety concerns in a cloud computing environment; data security, data transmission, virtual machine security, network security, data privacy, data integrity, data location, data availability, data segregation, compliances and patch management.

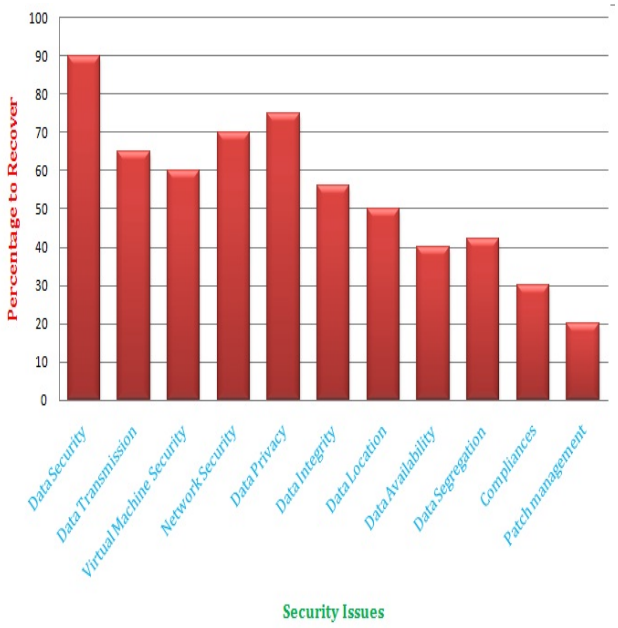

Figure 6. Big data security-key inhibitor of cloud computing

2 Related Work

In the past few years, a lot of research and development efforts have been made to define centralized and federated security mechanisms for the protection of identity information in a cloud environment. However, to the best of our knowledge none of the systems have been designed keeping anonymity as the key component. This paper describes an authentication and authorization protocol which outlines the main features of anonymous communication in the cloud [3]. In multi-level authentication cloud user secure access to the network and data centers. MLA protects cloud resources against unauthorized access by enforcing access control mechanisms. In our idea of MLA, We would like to outline our opinions about the usability of traditional text based password authentication along with biometrics authentication [4].

In this context different authentication schemes are implemented, such as multifactor authentication, access management, AWS identity. In this method the authentication that is allowing users to use just one password in order to authenticate themselves to multiple services. Authorization in the cloud computing is important for the users when they login to some cloud service because it enables prove of their identities. So, authorization is usually employed after the authentication. Oracle Database Vault is an example of security technique that enables authorization in the cloud. This security technique is offered by the vendor Oracle. Application data from different administrative users are protected with this authorization method [5].

Policy based authorization method that is protecting the privacy of the users enabling them to set privacy policies by themselves. In this way users are protecting their data in effective way from unauthorized access [6]. In this context authors proposed architecture of a hybrid cloud, composed of public cloud and private cloud. After data query from users, the private cloud store sensitive data after processing, and then send non sensitive data to public cloud. This architecture aims to achieve image data privacy via hybrid cloud, and reduce time of computation by dividing image on blocks and operate on these blocks, so that suitable for Big data [7].

An extension of big data technology framework is proposed by is based on security architecture. This security architecture is divided into the pre-filtering layer and the post-filtering layer. The pre-filtering layer is the first privacy layer of the proposed architecture. It finds and deletes personal sensitive information from the collected data and stores them in the matching database system database [8].

Big data comes with new trends for security and privacy. Big data analytics offer a large scale analysis and processing a huge amount of structured an uninstructed data in big data, for that becomes an important research area in security. Moreover, big data provide a great amount of information like data logs. [9].

Intelligent analysis platform for system construction of security log analysis using big data which is composed by collecting, saving, processing, and analyzing techniques. This architecture aims to analyze the relationship between security and data events created from network, system, application service of main IT infrastructure [10].

In other work authors outline that the user behaviours are dynamic which is difficult to capture the users' comprehensive behaviours in a single device by capturing or collecting the static dataset. Therefore, they proposed a log analysis system which is based on the Hadoop distribution platform to capture the traffic and analyze the user & machine behaviours, in terms of the search keywords, user shopping trends, website posts and replies, and web visited history to acquire the uses' dynamic behaviours [11].

A systematic framework for secure sensitive data sharing on a big data platform. This framework is composed of three components: security submission, security storage and security use. The basic flow of the framework is as follows: first personnel sensitive data are submitted to big data platform using security plug-in. After that, data stored in big data platform are encrypted with Proxy-re-encryption [12].

The features and technology offered by various providers created a great competitive market for the business. The various security issues are attracting attention, one of which is identity and privacy of the cloud user. Users are varied about the privacy of information which they have given to the provider at the time of registration [13].

Hence it would be advisable to have biometric authentication where members of house would be given access by configuring the system. But in order to enhance system's security it's important to have multiple hierarchy of authentication. Hence a PIN based authentication coupled with fingerprint recognition would enhance security of smart homes so that even if PIN is lost or stolen, a 2nd layer of authentication in the form of fingerprint would not give access to unauthorized users [14].

Programs written in this functional style are automatically parallelized and executed on a large cluster of commodity machines. The runtime system takes care of the details of partitioning the input data, scheduling the program's execution across a set of machines, handling machine failures, and managing the required inter-machine Communication. This allows programmers without any experience with parallel and distributed systems to easily utilize the resources of a large distributed system. Author proposes Simplified Data Processing on Large Clusters. Implementation of Map Reduce runs on a large cluster of commodity machines and is highly scalable: a typical Map Reduce computation processes many terabytes of data on thousands of machines [15].

Chen He Ying Lu David Swanson develops a new Map Reduce scheduling technique to enhance map task's data locality. He has integrated this technique into Hadoop default FIFO scheduler and Hadoop fair scheduler. To evaluate his technique, he compares not only Map Reduce scheduling algorithms with and without his technique but also with an existing data locality enhancement technique. Experimental results show that his technique often leads to the highest data locality rate and the lowest response time for map tasks. Furthermore, unlike the delay algorithm, it does not require an intricate parameter tuning process [16].

A prominent parallel data processing tool map reduce is gaining significant momentum from both industry and academia as the volume of data to analyze grows rapidly. While map reduce is used in many areas where massive data analysis is required, there are still debates on its performance, efficiency per node, and simple abstraction. This survey intends to assist the database and open source communities in under-standing various technical aspects of the map reduce frame work. In this survey, we characterize the map reduce frame work and discuss its inherent pros and cons [17].

Big data includes structured data, semi structured and unstructured data. Structured data are those data formatted for use in a database management system. Semi structured and unstructured data include all types of unformatted data including multimedia and social media content. Big data are also provided by myriad hardware objects, including sensors and actuators embedded in physical objects, which are termed the Internet of Things. Data storage techniques used for big data include multiple clustered network-attached storage (NAS) and object-based storage [18].

Password authentication is the commonly used single-factor authentication mechanism. The password authentication is defenceless to many security threats. Passwords are vulnerable to replay and discovery attacks. They also do not show any resistance to eavesdropping, man-in-the-middle or phishing attacks. Two-factor authentication opens up new horizons in security enhancement. It mandates users to provide two authentication tokens during the authentication phase. The two authentication tokens cover vulnerabilities of each other and combine together to provide higher information security [19].

User authentication is one of the most important parts of information security. Computer security most commonly depends on passwords to authenticate human users. Password authentication systems will be either been usable but not secure, or secure but not usable. While there are different types of authentication systems available alphanumeric password is the most commonly used authentication mechanism. But this method has significant drawbacks. An alternative solution to the text based authentication is Graphical User Authentication based on the fact that humans tend to remember images better than text. Graphical password authentication systems provide passwords which are easy to be created and remembered by the user [20].

3 Proposed Framework

In our proposed approach, we mainly focus on the user authentication, it involves user validation is a process that check and conclude a user's individuality. Authentication is the technique which is one of the most important factors in information assurance (IA). The remaining factors are availability, confidentiality, integrity and non repudiation. The authentication process starts when a client trying to use data in a secured environment. First and foremost, the client or user must show his right to use and individuality. When taking down into a workstation, clients generally go into client name and password for authentication process.

This login permutation, which has given to each and every user, validate right to use. Though, this category of certification can be circumvented by malicious user. Layer authentication is a data security (DS) administration method in which the uniqueness of a personality or organization is confirmed by two or

more validation procedure. It offers several levels of validation, depending on the essential operation, organization or working background. Layered verification is individuality and accurate to use management policies that are implemented in an environment that has an elevated exposure to vulnerability and rip-off. It is usually used to authenticate persons aforementioned to giving permission to a particular association and requires more than two proofs of individuality for authentication. Authentication is the key requirements to access the data in individual and organizations to promote a secure data transmission.

3.1 Merits of Multi Layer User Authentication

Increase flexibility and efficiency. The multi layer authentication reduces the burden of passwords by replacing them with alternatives. It has the potential to increase productivity and bring a better usability experience due to the increased flexibility of factor types.

Simplification of login process. It contains multiple layers and more advanced login options like distinct sign-on. The distinct sign-on validates the user during the login process. The unique sign on will covers the user's application, as what are all they needed.

Cyber Security. The multi-factor authentication (MFA) helps with cyber security because it is a combination of three or more authentication factors.

Improved protection. The multi factor authentication can be used to provide added security. It holds the additional protection in each layer.

Attain compliance. Another benefit of multi factor authentication is being able to achieve the necessary compliance requirements specific to the organization.

The following Table 1 shows the various security methods are used in the multifactor authentication. In multifactor authentication three security mechanisms are using, the username and password of the customer is verifying , the onetime password has been sent to customer for verification and biometric information are verified.

Table 1. Security methods

Security Methods / Authentication	User name & password verification	OTP Verify cation	Bio metric verify cation
Single factor	√	-	-
Two factor	√	√	-
Multi factor	√	√	√

3.2 Proposed Multi Layer User Authentication

The proposed approach follows the authentication process in following way. The data stored in the cloud can be grouped into various categories based on their usage. It is grouped as like most frequently used, frequently used and rarely used. Here, we provide the multi factor authentication technique to offer the easy access and to secure the data; it contains the multiple layers of authentication. In first layer, the user needs to enter their user credentials and password. After verifying that, the user needs to move next authentication. In second authentication, the user has to provide onetime password with his registered identity. In this level, the user needs to enter the onetime password. After authenticating the onetime password, the user allowed to move for the next layer. The third authentication is the important one; here the user needs to deliver the credentials about the data that they are trying to access and then their biometric identity. This authentication provides the ease access of data and increased security. By using the third authentication, we can have the data integrity, confidentiality, availability and security.

The above Figure 7 shows that the authentication of conceptual methods. In proposed structure cluster information centers split into a series of m data units, every cluster data center noted by a (1, m), and the information units are placed at n dissimilar storage space providers, where each provider is identified as b (1, n). In common, (units of data center) m is for all time larger than (number of service provider) n, the n storage providers are belongs to dissimilar organizations such as IBM, VMware, Microsoft, AWS and Google. Proposed system provides three different layers of authentication for accessing a big data in cloud environment.

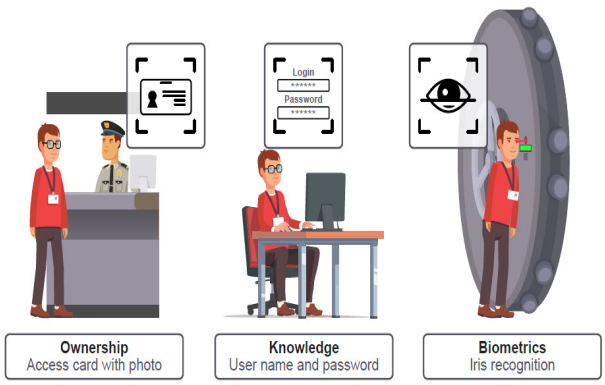

Figure 7. Conceptual authentication

Layer 1 Authentication: In this authentication process the service provider verify the username and password of the user. If the username and password is matched with registered details, then service provider give authentication to particular user for entering next layer. The login details are maintained in the login table by the service provider. If the user is not authenticated by the service provider, then the unauthorized account holder details are stored in the malicious table.

Algorithm 1. Username and Password Authentication

Input: n= total number of username
m= total number of password
Target: target authentication
Output: authenticated user

1. Begin
2. pair for j = 1 to n
3. pair for k = 1 to m
4. if (j == registered user)
5. if (k == registered password)
6. then
7. user = authorized
8. user go to next authentication
9. else
10. user = unauthorized
11. entry to table
12. End

The above Figure 8 shows the single factor authentication can be developed in to two factor authentication with some security features. The two factor authentication is further developed and has high security features. The multifactor authentication is a most advanced authentication technique in big data security.

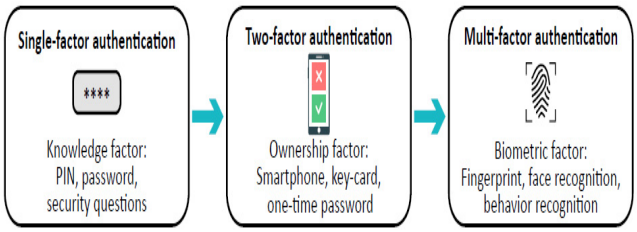

Figure 8. Evolution of authentication methods

Layer 2 Authentication: After successfully completing the first authentication process, the service provider sending a onetime password message will be sent to user registered mobile phone. After successfully verifying the onetime password, the service provider directs the users to enter next level authentication process. The authentication details are stored in the authorization table, if any malicious users trying to access the data center, these details are stored in malicious user table.

Algorithm 2. One Time Password Verification

Input: N= Registered user mobile number
Target: Mobile number verification
Output: Authorized user

1. Begin
2. i = mobile number
3. if (i = = registered mobile number)
4. then
5. user = authorized
6. user go to next authentication
7. else
8. user = unauthorized
9. entry to table
10. End

Layer 3 Authentication: This is the last authentication method, in this user going to access the data unit. While accessing data unit user need to get authenticate. The biometric (finger print or retinal) detail of the user is verified here, after authentication processes over at last the user is allowed to access the data. Based on their request the corresponding data canters will be directed. The authenticated user's details are stored in the authentication table and the unauthorized user's details are stored in the malicious user table.

Algorithm 3. Bio – metric Authentication

Input A: user photo image
Target: matching with image
Output: allow user to access data

1. Begin
2. A = input image
3. if (A == DB image)
4. then
5. user = authenticated
6. authenticated user access data unit
7. else
8. user = unauthorized
9. entry to table
10. End

The large information is separating and placed in various data center; supervisor of the complete organization maintain storeroom space rate for every information units. Client accessing the application in scattered cloud system are explained in Figure. 8. The centralized cluster maintaining the details about the users, who are all accessed the data center in distributed environment. Proposed algorithm maintains the subsequent database (1) Malicious user data (2) Authorized user information (3) IBM cloud data storage data (4) VMware data storage. Malicious user table will store the access associated to unauthorized effort, most frequently used, frequently used – rarely used records are placed in another data base.

The above Figure 9 shows the flow of client accessing the applications in distributed system. Client sending a query message to centralized cluster data center for accessing application. The Centralized data unit processing the client request and direct the client for corresponding data center. Based on the client request centralized data unit direct the client to most frequently used data unit or frequently used data center or rarely used data center.

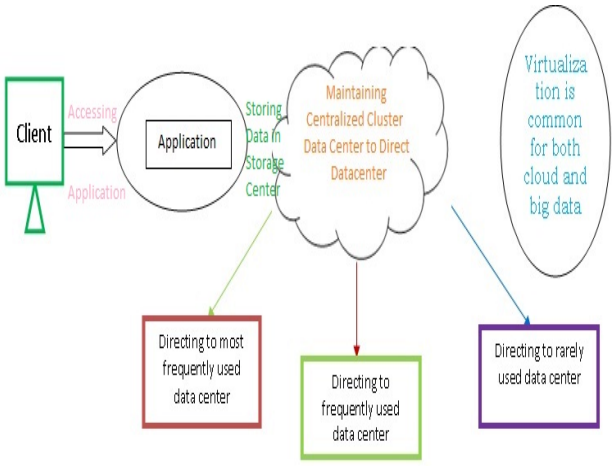

Figure 9. Client accessing application in distributed cloud system

/*Route the assessment log folder to confirm the logged in user's details like his/her/device path location, method date etc.*/

Select username, password, registered mobile, user image, last-logged in, present position, geolocation, ipconfig, method date;

```
//If some unconstitutional consumer tried to entrance request
{
Put in into hateful reorganized values
(Username, pass code, geolocation, method date);
return -1;
}

else
{
List A = Select Storage provider, data center from cluster storage Cloud where Customer_loggedin_ ApplicationId=? ;

If (A. Storage provider == "IBM Cloud" && A. data center == "most frequently used")
{

Goto IBM Cloud data storage
Select * from IBM cloud data storage where customer_loggedin_ApplicationId=? ;
}

else if (A. Storage provider == "VMware cloud" && A. data center == "frequently used")
{

Goto VMware Cloud data storage
Select * from VMware cloud data storage where customer_loggedin_ApplicationId=? ;
}

else if (A. Storage provider == "Microsoft cloud"&& A. data center == "rarely used")
{
Goto Microsoft cloud data storage;
Select * from Microsoft cloud data storage where client_loggedin_ApplicationId=? ;
}
}
}//end else
```

$$\sum Au = £tu - £mu$$

Where:

$\sum Au$ - No. of Authorized users

£tu - Total no. of users logged in

£mu - Malicious no. of users

3.3 Security for Cluster Data in Cloud

Customer accessing application in cloud, they need to prove them as a valid user by providing some personal identities which is already registered with their service provider. We have various cloud service providers like IBM, VMware and Microsoft. Figure.9 shows that the layer authentication of IBM cloud service provider. This architecture diagram clearly tells the layer authentication process.

The proposed system uses a shared security system for securing the data in distributed cloud system. The IBM cloud is divided in to three clusters based on the usage of data, most frequently used, frequently used and rarely used. First the user need to get authenticate in layer1, the layer 1 authentication provide security for IBM cloud. In this, customers need to authenticate with their credentials like user name and password.

Table 2. Authentication techniques

Authentication / Security Improvement	Single factor	Two factor	Multi factor
Availability	Low	Medium	High
Data Security	Low	Medium	High
Malicious Users	High	Low	Very Low
Data Protection	Low	High	Very High

Table 2 shows the improvement of security when we use multi factor authentication. The multi factor authentications improve the data protection in very high manner similarly the malicious users were reduced very low. Availability of the data rate gets increased when compared with single factor and two factors, and also security of the data is high. The multifactor provides the efficient data security and effective data protection.

3.4 Big Data Security in Distributed Cloud Computing Environment

In distributed environment, big data security and privacy protection are the two main factors of user's concerns about the cloud technology. Here we are using the IBM cloud service provider to store the data. Big data is a term that describes the large volume of data; the data may be structured or unstructured.

The importance of big data doesn't revolve around how much data you have, but what you do with it. When big data is stored in distributed network, there is a huge possibility for security frightening. We propose a multi factor authentication technique to secure the big data in distributed network computing. In each layer, we use different authentication mechanism's (something you know, something you have, and something you are (biometrics)) to protect a big data from the hacker.

A layer 2 authentication is providing for cluster data center, once the users are authenticated in layer1, they have directed to access the cluster data centers. In this authentication customer need to verify their one time password using already registered mobile number, after verifying the one time password of the customer they need to go for layer 3 authentications, the unauthorized user's detail are entered in entry table.

In layer 3 authentication the individual data units are accessed by the users, for accessing the data unit user need to provide their photo image or finger print for authentication process. The following Figure 10 shows the security architecture of the cluster data storage.

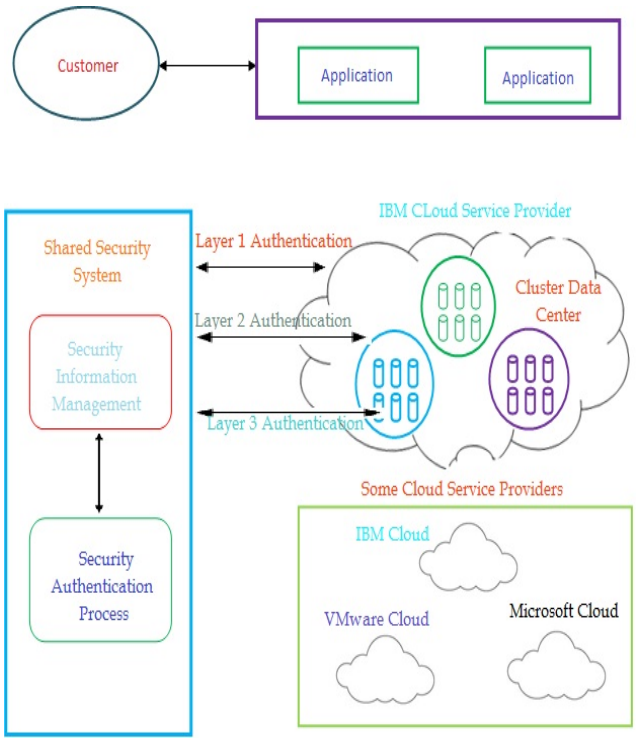

Figure 10. Security architecture for cluster based data storage in cloud

3.5 Dispensation of Large Information in AWS

Apache Hive is overprotective unfasten establishment software that runs on peak of Hadoop in Amazon Elastic Map Reduce (AEMR). There is no particular software or procedures necessary to generate an interactive Hive conference in amazon elastic map reduce (AEMR) webpage. Generating a communicative hive conference does not need any ladder to be additional or configured in elastic map reduce (EMR) web page. In universal, Hive and Pig are introduced by defaulting on each new bunch of Amazon EMR. One time the user effectively linked with main lump in the cluster, and then it is simple to call upon the hive instructed in a straight line on the EMR cluster. It offers SQL-like commands to right of entry the data. In this anticipated architecture Hive is used to method the log records saved in the Amazon S3 [21].

3.6 Map Reduce for Dispensation Record Information in Spotted Centre

Map Reduce is a curriculum representation or construction that progression everyday jobs in equivalent transversely a huge quantity of systems. It can have two methods such as Mapping and Reducing. Mapping method divide the enormous amount of input data addicted to <key, value> pairs. Intermediate <key, value> pairs will be produced based on aggregating quite a few input key value pairs from the Mapping point [22].

Ultimately, reducing phase takes the intermediary key rate pairs and generate the productivity <key, value> pairs with the intention can be simply understand by the closing stages customer in this planned structural design, map reduce skeleton is used to discover the amount of clients who have logged in to the cloud information centre. Projected map reduce simulated code can powerfully progression the enormous amount of log folder in which it can have users who were logged in with date and the register in instance interval [22].

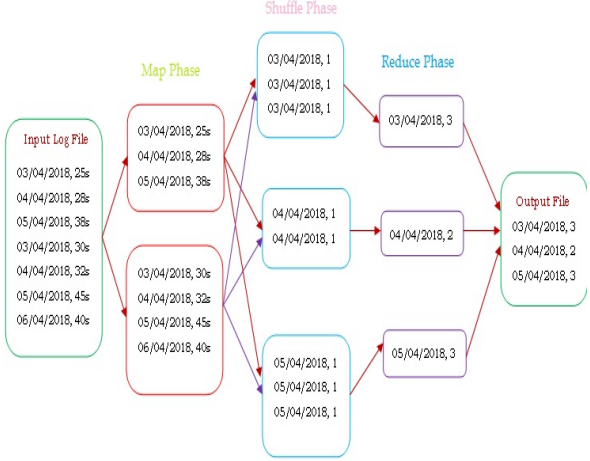

Figure 11. Skeleton for processing log files

The following Figure 11 shows the skeleton of log file processing.

```
class MAPPER
    method MAP(docid a, doc d)
        for all term t ∈ doc d do
            EMIT(term t, count 1)
class REDUCER
    method REDUCE(term t, counts [c1, c2, ...])
        sum ← 0
        for all count c ∈ counts [c1, c2, ...] do
            sum ← sum + c
        EMIT(term t, count sum)
```

The following Figure.12 shows that the list of service providers with the quantity of data centers, The IBM cloud service provider will have one thousand data centers and five thousand data units. In VMware cloud seven hundred and fifty data centers and three thousand two hundred data units. In Google cloud it has one thousand four hundred and fifty data centers and six thousand nine hundred and fifty data units.

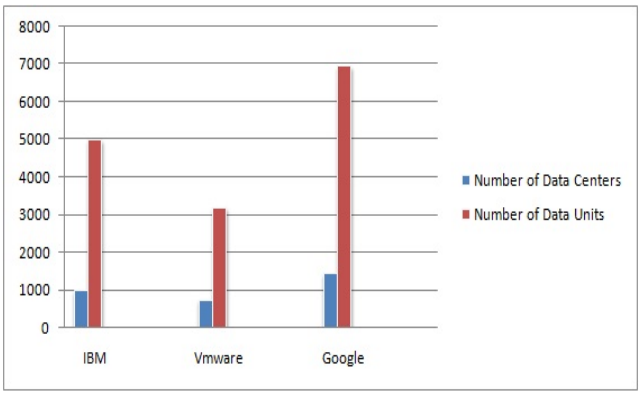

Figure 12. Service providers with cluster data centre

The above Figure.13 shows that the detail of total number of users logged in on a particular day, it also shows that the numbers of authorized users were logged in and the numbers of malicious users were logged in. Every day the above details are updated regularly.

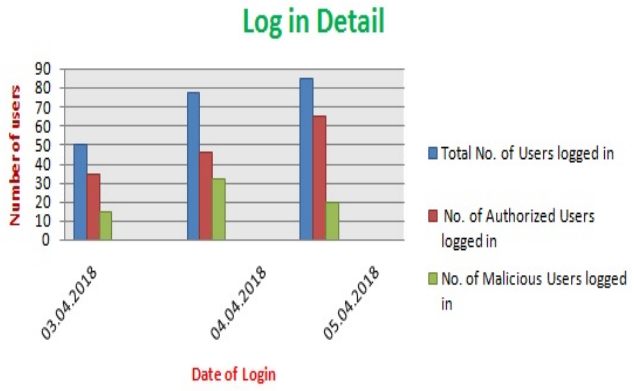

Figure 13. User logged in detail

A comparison of the main indicators is already deployed and emerging factors is given in Table 3. The factors are evaluated based on the following parameters: Universality stands for the presence of factor in each person; Uniqueness indicates how well the factor differentiates one person from another; Collectability measures how easy it is to acquire data for processing; Performance indicates the achievable accuracy, speed, and robustness; Acceptability stands for the degree of acceptance of the technology by people in their daily life; Spoofing indicates the level of difficulty to capture and spoof the sample. However, many other issues are to be addressed while integrating the MFA for the end users. In the following section, we elaborate on those challenges and formalize the recommendations for improved ease of integration [30].

Table 3. Comparison of suitable factors for MFA: H-High; M-Medium; L-Low; N/A-Unavailable

Factor	Universality	Uniqueness	Collectability	Performance	Acceptability	Spoofing
Password	n/a	L	H	H	H	H
Token	n/a	M	H	H	H	H
Voice	M	L	M	L	H	H
Facial	H	L	M	L	H	M
Fingerprint	M	H	M	H	M	H
Hand geometry	M	M	M	M	M	M
Location	n/a	L	M	H	M	H

The above Figure 14 shows that the comparison of existing two factors and three factor authentication with our system. When compared with previous techniques, our cluster based proposed approach have the high impact in security level, information manageable, data control, governance, security risk and authentication biometrics. Our authentication technique provides the efficient data transformation in distributed cloud environment.

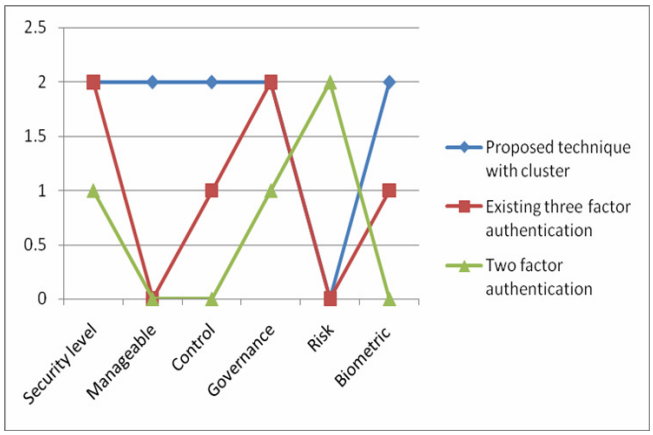

Figure 14. Comparison of various factor authentications

The above Table 4 shows that the comparison of different authentication schemes, the authentication parameters are categorized in to five C's. It denotes, user can easily choose their password, anonymity of the user, user can able to change the password securely, multifactor authentication and mutual authentication between service provider and users. Our proposed scheme satisfy all the five C's and produce high security authentication for data stored in the cloud. Compared with existing systems, the proposed system attain high secured authenticated data storage in distributed cloud computing.

Table 4. Comparison of authentication schemes

	Our Scheme	Das et al. [27]	Chien et al. [28]	Pathan et al. [29]
C1	Yes	Yes	Yes	Yes
C2	Yes	Yes	No	No
C3	Yes	No	No	No
C4	Yes	No	No	No
C5	Yes	No	Yes	Yes

Note. C1: Freely chosen password; C2: User anonymity; C3: Secure password change; C4: Multifactor security; C5: Mutual authentication.

In above Figure 15, we compared various data accessibility methods with our proposed scheme. Our proposed scheme has high accessibility of information, accuracy of the data and availability of the information; our clustering data centre direct the users to access individual data units based on their request. Our proposed scheme makes easy to access the data with high security authentication.

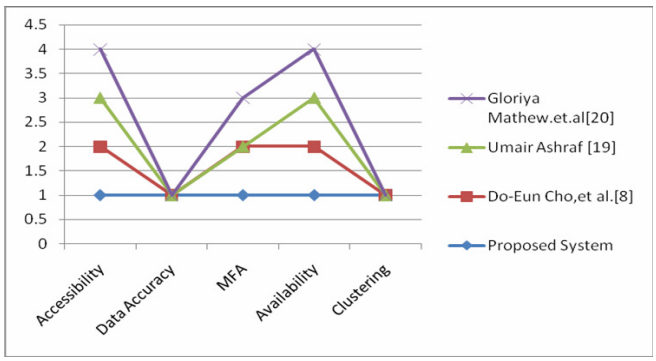

Figure 15. Comparison of data accessibility

4 Conclusion and Future Enhancement

In proposed system we developed narrative cluster based multilayer user authentication information storeroom architecture for securing big data from malicious users in distributed cloud computing environment. User is authenticating in each layer using different authentication techniques. Compared with existing two factors and three factor authentication, the proposed cluster based multifactor or multilayer authentication provides more security, high authentication and ease of access information in cloud. Clustering the data centre is an added advantage of our proposed authentication system. Chart reduce skeleton is used to discover the amount of users who are all login into the cloud information unit. The anticipated structure provides protection for the mapping of a variety of information essentials to every supplier using cluster based multifactor authentication data storage boundary. This projected approach requires eminent implementation attempt; it provides significant data for cloud environment conditions that can be capable of have elevated collision on the subsequently innovation systems. When the user using this multilayer authentication technique to transfer the data, they really feel that their data's are secure. Our upcoming effort is to make longer the multifactor authentication data storage in to four factor authentication with digital signature.

References

[1] P. Mell, T. Granceet, *The NIST Definition of Cloud Computing, Reports on Computer Systems Technology*, NIST Special Publication 800-145, September, 2011.

[2] A. Katal, M. Wazid, R. H. Goudar, Big Data: Issues, Challenges, Tools And Good Practices, *2013 Sixth International Conference on Contemporary Computing*, Noida, India, 2013, pp. 404-409.

[3] U. Khalid, A. Ghafoor, M. Irum, M. A. Shibli, Cloud Based Secure and Privacy Enhanced Authentication & Authorization Protocol, *Procedia Computer Science*, Vol. 22, pp. 680-688, 2013, DOI: 10.1016/j.procs.2013.09.149.

[4] S. Ahmad, B. Ehsan, The Cloud Computing Security Secure User Authentication Technique, *International Journal of Scientific & Engineering Research*, Vol. 4, No. 12, pp. 2166-2171, December, 2013.

[5] K. Jakimoski, Security Techniques for Data Protection in Cloud Computing, *International Journal of Grid and Distributed Computing*, Vol. 9, No. 1, pp. 49-56, 2016.

[6] D. W. Chadwick, K. Fatema, A Privacy Preserving Authorisation System for the Cloud, *Journal of Computer and System Sciences*, Vol. 78, No. 5, pp. 1359-1373, September, 2012.

[7] K. Shirudkar, D. Motwani, Big-Data Security, *International Journal of Advanced Research in Computer Science and Software Engineering*, Vol. 5, No. 3, pp. 1100-1109, March, 2015.

[8] D.-E. Cho, S. J. Kim, S.-S. Yeo, Double Privacy Layer Architecture for Big Data Framework, *International Journal of Software Engineering and Its Applications*, Vol. 10, No. 2, pp. 271-278, February, 2016.

[9] A. A. Cardenas, P. K. Manadhata, S. P. Rajan, Big Data Analytics for Security, *IEEE Security & Privacy*, Vol. 11, No. 6, pp. 74-76, November-December, 2013.

[10] K.-S. Jeon, S.-J. Park, S.-H. Chun, J.-B. Kim, A Study on the Big Data Log Analysis for Security, *International Journal of*

Security and Its Applications, Vol. 10, No. 1, pp. 13-20, January, 2016.

[11] A. Kourid, S. Chikhi, S.-P. Hong, A Comparative Study of Recent Advances in Big Data Security and Privacy, *Asia-pacific Journal of Multimedia Services Convergent with Art, Humanities, and Sociology*, Vol. 7, No. 5, pp. 873-883, May, 2017.

[12] X. Dong, R. Li, H. He, W. Zhou, Z. Xue, H. Wu, Secure Sensitive Data Sharing on a Big Data Platform, *Tsinghua Science and Technology*, Vol. 20, No. 1, pp. 72-80, February, 2015.

[13] R. Shaikh, M. Sasikumar, Identity Management in Cloud Computing, *International Journal of Computer Applications*, Vol. 63, No. 11, pp. 17-19, February, 2013

[14] P. S.-P. Wang, S. N. Yanushkevich, Biometric Technologies and Applications, *AIAP'07 Proceedings of the 25th Conference on Proceedings of the 25th IASTED International Mult –Conference: Artificial Intelligence and Applications*, Innsbruck, Austria, 2007, pp. 226-231.

[15] J. Dean, S. Ghemawat, MapReduce: Simplified Data Processing on Large Clusters, *In Proceedings of Operating Systems Design and Implementation*, San Francisco, CA, 2004, pp. 137-150.

[16] C. He, Y. Lu, D. Swanson, Matchmaking: A New MapReduce Scheduling Technique, *2011 IEEE Third International Conference on Cloud Computing Technology and Science*, Athens, Greece, 2011, pp. 40-47.

[17] K.-H. Lee, Y.-J. Lee, H. Choi, Y. D. Chung, B. Moon, Parallel Data Processing with MapReduce: A Survey, *Special Interest Group on Management of Data Record*, Vol. 40, No. 4, pp. 11-20, December, 2011.

[18] B. Purcell, The Emergence of "Big Data" Technology and Analytics, *Journal of Technology Research*, Vol. 4, pp. 1-6, July, 2013.

[19] U. Ashraf, *Securing Cloud Applications with Two-Factor Authentication*, Master Thesis, University of Stuttgart, Stuttgart, Germany, 2013.

[20] G. Mathew, S. Thomas, A Novel Multifactor Authentication System Ensuring Usability and Security, *International Journal of Security, Privacy and Trust Management*, Vol. 2, No. 5, pp. 21-30, October, 2013.

[21] K. Schmidt, C. Phillips, *Programming Elastic MapReduce: Using AWS Services to Build an End-to-end Application*, O'Reilly Media, 2013.

[22] G. Manogaran, C. Thotab, M. V. Kumar, MetaCloud DataStorage Architecture for Big Data Security in Cloud Computing, *Procedia Computer Science*, Vol. 87, pp. 128-133, December, 2016.

[23] Wikipedia, *Big Data*, https://en.wikipedia.org/wiki/Big_data

[24] H. S. Lamba, G. Singh, Cloud Computing-Future Framework for e-management of NGO's, *International Journal of Advancements in Technology*, Vol. 2, No. 3, pp. 400-407, July, 2011.

[25] M. Ahmed, M. A. Hossain, Cloud Computing and Security Issues in the Cloud, *International Journal of Network Security & Its Applications*, Vol. 6, No. 1, pp. 25-36, January, 2014.

[26] G. Singh, S. Sood, A. Sharma, CM- Measurement Facets for Cloud Performance, *International Journal of Computer Applications*, Vol. 23, No. 3, pp. 37-42, June, 2011.

[27] M. L. Das, A. Saxena, V. P. Gulati, A Dynamic ID-based Remote User Authentication Scheme, *IEEE Transactions on Consumer Electronics*, Vol. 50, No. 2, pp. 629-631, May, 2004.

[28] H. Y. Chien, J. K. Jan, Y. M. Tseng, An Efficient and Practical Solution to Remote Authentication: Smart Card, *Computers and Security*, Vol. 21, No. 4, pp. 372-375, August, 2002.

[29] A. K. Pathan, C. S. Hong, T. Suda, A Novel and Efficient Bilateral Remote User Authentication Scheme Using Smart Cards, *2007 Digest of Technical Papers International Conference on Consumer Electronics*, Las Vegas, NV, USA, 2007, pp. 1-2.

[30] W.-Y. Yau, The "123" of Biometric Technology, *Synthesis Journal*, Section 3, pp. 83-96, 2002.

Biographies

S. Ramasamy, received B.Tech degree from M. Kumarasamy College of Engineering, Tamilnadu, India and M.E. degree from Mahendra Engineering College. He is pursuing Ph.D. degree in ICE at Anna University, Chennai. His current research interests include cloud computing security issues and big data security analysis.

R. K. Gnanamurthy received B.E degree from Bharathiar University and M.E. degree in Madurai Kamaraj University and Ph.D. degree in ICE from Anna University, Chennai. He has authored more than 15 publications in journals and 18 publications in conferences. His research interests include Image Processing, Wireless Sensor Networks, Mobile Computing, Cluster and Cloud Computing.