An Improvement In Cloud Data Security That Uses Data Mining

Prof.P. S. Tembhurne,Prof.C.M.Goswami,Prof. S.V.Deshmukh

Department of computer science &egg, Prof.Ram Meghe institute of technology and research,Badnera,India

*Abstract:**Cloud Computing provides a good model for the providers to deploy the computing infrastructure and applications on-demand. It offers greater flexibility to users by connecting to various computing resources and allowing access to IT enabled services. But it has the risk of privacy of user data and security. Despite its growing influence, concerns regarding cloud computing still remain. In our opinion, the benefits outweigh the drawbacks and the model is worth exploring. Some common challenges are: data protection.data recovery and availability. Cloud Computing Environment Should secure enough in maintaining cloud users trust.Secure client data & Communication requires for evaluating cloud security. Identifying unique security requirements for feasible solution that eliminates potential threats, it provides secure inter-working for maintaining the confidentiality integrity of information. In this Paper I have ReviewFor Static Data Security: AESfor Dynamic Data Security: RSA,Communication Security Error Processing: Errors are detected during fault treatment while accessing Checksum algorithm.*

Keyword:AES(Advanced Encryption standard)RSA(RivestShamir,adleman)

I. INTRODUCTION

The name cloud computing was inspired by the cloud symbol that's often used to represent the Internet in flowcharts and diagrams Cloud computing is the service Offer to various end users, companies.Cloud computing enables the end-users, small and medium-sized companies to access computational resources like storage, software etc. In cloud computing, with these vast amount of computing resources, users are able to solve their problems easily using the resources provided by cloudUsing the cloud computing services, users are able to store their data in servers and access their data from anywhere and they need not worry about the lose of data due to disk faults, system breakdown etc. But there are several security issues in cloud like assurance and confidentiality of user data. The users who are entrusting the cloud provider may lose the access to his data either permanently or temporarily due to any unexpected event like malware attack. This unexpected event provides significant harm to the users. The providers in cloud can analyze the user data continuously and similarly the outside attackers who try to get access to the cloud can also analyze the user data. So, the user may lose his data privacy.Classification is a process where sensitive data is identified and appropriate mechanisms are implemented to maintain privacy of this sensitive data. Fragmentation is a process where the data is divided into small chunks. Distribution is a process where the divided chunks will be distributed to cloud providers. Distribution of data to a cloud provider can be done depending upon the reliability of cloud provider and data sensitivity. The reliability of a cloud provider means if the cloud provider is able to store the data chunks with such sensitivity.

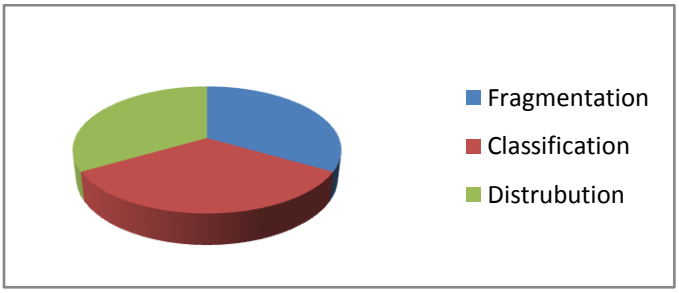

Fig 1: Showing process of Identifying Data

Using this approach, it is difficult for the attacker to get the data chunks from different providers and also mining sensitive information from these data chunks is a tedious process. this paper explain defense from attack. This deals with the front end and the back end. The front end is the side the computer user, or client, sees. The back end of the system is the various computers, servers and data storage systems that create the "cloud" of computing services Cloud computing will need to find ways to protect client privacy Cloud has centralized server administration system Centralized server administers the system, balances client supply, adjusts demands, monitors traffic.

Cloud Computing Environment Should secure enough in maintaining cloud users trust.secure client data & Communication requires for evaluating cloud security. Identifying unique security requirements for feasible solution that eliminates potential threats, it provides secure inter-working for maintaining the confidentiality integrity of information. Security requirements are Authentication, Authorization, and Confidentiality.

If a client can log in from any location to access data and applications, it's possible the client's privacy could be compromised. Cloud computing will need to find ways to protect client privacy. One way is to use authentication techniques such as user names and passwords & cryptography which is important and can be used as reference for designing the complete security solution.

ISSN: 2278 – 1323 All Rights Reserved © 2015 IJARCET

CLOUD STANDARD SERVICES:

- Software as a Service (SAAS):We use the provider apps, User doesn't manage or control the network, servers, OS, storage or applications
- Platform as a Service (PAAS): User deploys their apps on the cloud, Controls their apps, User doesn't manage servers, IP, storage
- Infrastructure as a Service (IAAS): Consumers gets access to the infrastructure to deploy their stuff, manage or control the infrastructure, Does manage or control the OS, storage, apps, selected network

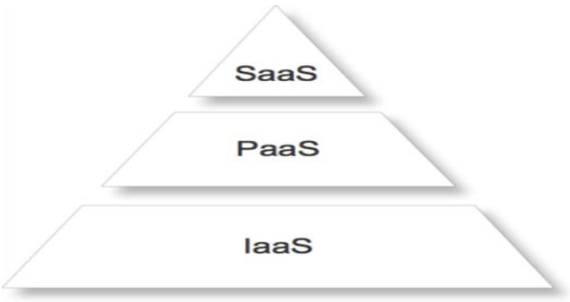

Fig2: Standard Services

In this paper, we present a framework for security in cloud computing, data security, providing availability, integrity, confidentiality. This cloud does not account some essential characteristic of cloud computing scalability, elasticity, computer virtualization, relativeconsistency, reliability. This Paperserves data storage.
Service & Providing static & dynamic security and making cloud services fault tolerant, secure.

Feasibility of this project, under IAAS, storage as a service is secured enough and performance associates with checking each data which is stored at datacenters clone copy of that on datacenters replicated by cloud-coordinator. Before sending file/data on cloud that data encrypted with AES (static data) & transit data enveloped with RSA [3].
Checksum Class is used for to providers fault tolerant data servers on cloud, each bit of data on datacenters we can check with the help of checksum class.
Analysing allabove thing Paper workfocuseson"Reviewof data security in cloud computing".This paper also help to aware about efficiency against malicious data modification attacks.

II. BACKGROUND&LITERATURE SURVEY

A]Problem Statement:
A survey conducted by IDC (International Data Corporation) suggests that cloud services are still in the early adoption phase.The survey has rated security as the most prominent concern.

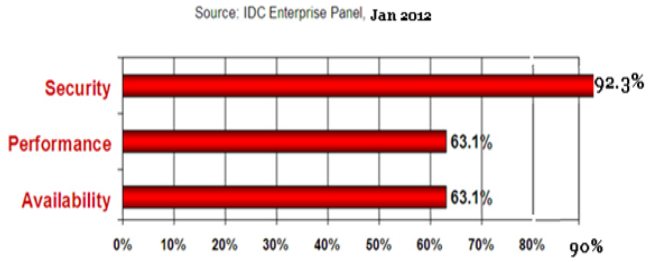

Fig 3:Survey of IDC in 2012

Hence In this paper we try to overcome this concern
By proposing for data security:AES, for data transmission:RSA for Error recovery :MD5 algorithm.

B. *Origin and Definition of Cloud Computing*
The Internet rapidly began to grow up in the 1990s and, the progressively more complicated network infrastructure and enlarged bandwidth developed in the recent years have considerably improved the strength of various application services available to users through the Internet, hence, marking the beginning of cloud computing network services. Cloud computing services use the Internet as a communication medium and convert information technology resources into services for end-users, including software services, computing platform services, development platform services, and basic infrastructure leasing.

Primary significance of Cloud computing lies in allowing the end users to access computation resources through the Internet. The unusual features of cloud computing include the storage of user data in the cloud and the lack of any need for software installation on the client side. Provided that the user is able to connect to the Internet, all of the hardware resources in the cloud can be used as client-side infrastructure.Normally, cloud computing applications are demand-driven, providing various services according to user requirements, and service providers charge by metered time, instances of use, or defined period[1]. Cloud computing can be defined as "a type of parallel and distributed system which consists of a collection of interconnected and virtualized computers that are dynamically provisioned and presented as one or more unified computing resources based on service-level agreements established through negotiation between the service provider and consumers"[2]. The cloud computing concept can be understood in a more better way by following the below given figure.

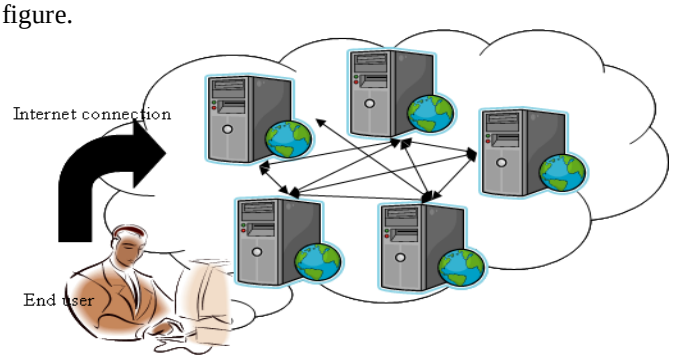

Fig 4: Pictorial Representation of Cloud computing

ISSN: 2278 – 1323 All Rights Reserved © 2015 IJARCET

C.Origin and Definition of Data Mining

Data mining is to find knowledge, and knowledge is represented through certain patterns. Association rule is the most often used method in data mining, which finds out the association between data and various objects by finding the potential dependence among data. Classification and clustering can be used to sort out things by characterizing the common significance among different things. The disadvantage of data mining in centralized database, generally have the several following points: network traffic is considered less, mining efficiency is low and the degree of spatial complexity is high. The most classic classification data mining are classification methods based on distance, classification methods based on decision tree, Bayesian classification and so on. Data mining techniques have been extensively used in various applications. However, the mistreat of these techniques may lead to the discovery of sensitive information. Researchers have recently made efforts at hiding sensitive association rules. However, undesired side effects, e.g., non-sensitive rules falsely hidden and spurious rules falsely generated, may be formed in the rule hiding process. [3]

The following figure explains the different steps which comprise the overall data mining process:

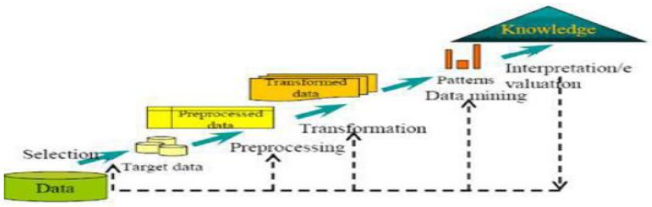

Fig 5: Pictorial Representation of Data Mining.

Following Are some Ealier works on this mehods:

a.Sussman and Booz Allen alumnus James Hutchinson came up with the concept of using cloud computing—Internet-based applications to drive up the performance of these large biometric databases, Biometric technology that promises to be limitless in the amounts of data it can analyze and to distribute data rapidly to anyone, anywhere thousands of records of biometric data, and run it on a cloud computing platform to document expected gains from implementing cloud computing solutions[4].

b. Lilli bridge et al. presented a P2P backup scheme in which blocks of a data file are dispersed across m+k peers using an (m+k,m)-erasure code. Peers can request random blocks from their backup peers and verify the integrity using separate keyed cryptographic hashes attached on each block. Their scheme can detect data loss from free riding peers, but does not ensure all data is unchanged[5].

c. Shah et al proposed allowing a TPA to keep online storage honest by first encrypting the data then sending a number of pre-computed symmetric-keyed hashes over the encrypted data to the auditor. However, their scheme only works for encrypted files and auditors must maintain long-term state[6].

d.Schwarz et al. proposed to ensure file integrity across multiple distributed servers, using erasure-coding and block-level file integrity checks. However, their schemes only considers static data files and do not explicitly study the problem of data error localization[7].

III THE ALGORITHM TO BE USED.

A. ***AES for Static Data Security:***

National Security Agency (NSA) strength of all key lengths of the AES algorithm (i.e., 128, 192 and 256) is **sufficient** to **protect** classified information up to the **SECRET** level. **TOP SECRET** information will require use of either the 192 or 256 keylength[8].

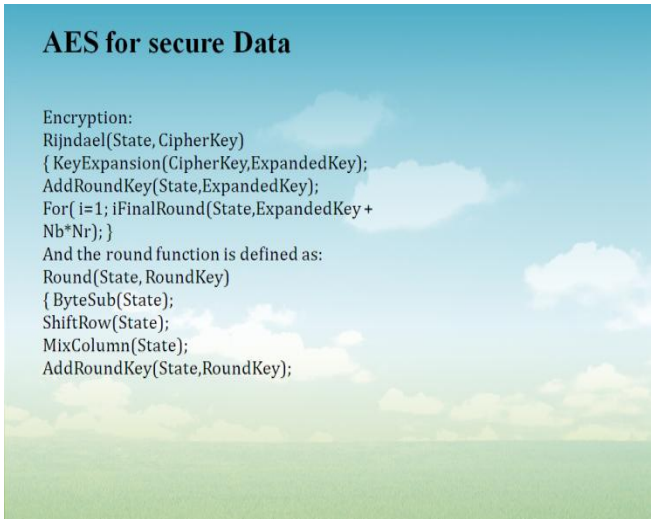

Fig 6: AES In General

1. In the Sub Bytes step, each byte in the state is replaced with its entry in a fixed 8-bit lookup table, *S*; $b_{ij} = S(a_{ij})$.

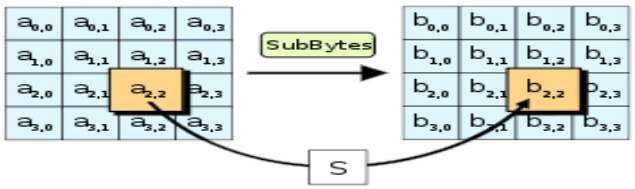

Fig 7: Subyte transformation

2. In the Shift Rows step, bytes in each row of the state are shifted cyclically to the left. The number of places each byte is shifted differs for each row.

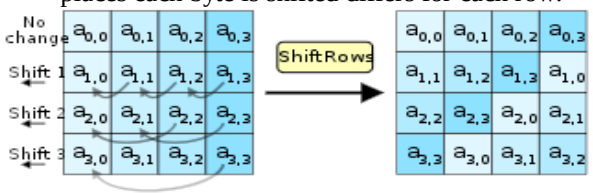

Fig 8: Shift rows

ISSN: 2278 – 1323 All Rights Reserved © 2015 IJARCET

3. In the Mix Columns step, each column of the state is multiplied with a fixed polynomial *c(x)*.

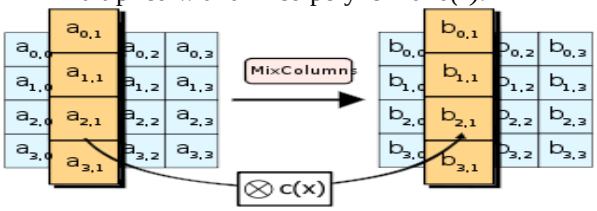

Fig 9: Mix Column

4. In the AddRoundKey step, each byte of the state is combined with a byte of the round sub key using theXOR operation ($\oplus$)

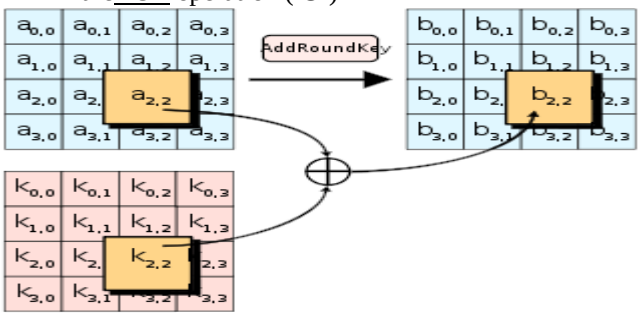

Fig 10: Add round key

AES operates on a 4×4 column-major order matrix of bytes, termed the *state*, although some versions of Rijndael [13] have a larger block size and have additional columns in the state. Most AES calculations are done in a special finite field. The key size used for an AES cipher specifies the number of repetitions of transformation rounds that convert the input, called the plaintext, into the final output, called the ciphertext. The numbers of cycles of repetitions are as follows:

- 10 cycles of repetition for 128-bit keys.
- 12 cycles of repetition for 192-bit keys.
- 14 cycles of repetition for 256-bit keys.
- Hence the cipher text c = 13, to check decryption we compute m' = c^d mod n = 13^7 mod 33 = 7.

- Note that we don't have to calculate the full value of 13 to the power 7 here. We can make use of the fact that a = bc mod n = (b mod n).(c mod n) mod n so we can break down a potentially large number into its components and combine the results of easier, smaller calculations to calculate the final value. One way of calculating m' is as follows:-

- m' = 13^7 mod 33 = $13^{(3+3+1)}$ mod 33 = $13^3.13^3.13$ mod 33= (13^3 mod 33).(13^3 mod 33).(13 mod 33) mod 33= (2197 mod 33).(2197 mod 33).(13 mod 33) mod 33= 19.19.13 mod 33 = 4693 mod 33= 7

Each round consists of several processing steps, including one that depends on the encryptionkey itself. A set of reverse rounds are applied to transform cipher text back into the original plaintext using the same encryption key. These are the following steps (AES Processing Step) as shown in figure:

1. Key Expansion—round keys are derived from the cipher key using Rijndael's key schedule [13].
2. Initial Round
   a. `AddRoundKey`—each byte of the state is combined with the round key using bitwise xor.
3. Rounds
   a. SubBytes: anon-linear substitution step where each byte is replaced with another according to a lookup table.
   b. ShiftRows: a transposition step where each row of the state is shifted cyclically a certain number of steps.
   c. MixColumns: a mixing operation which operates on the columns of the state, combining the four bytes in each column.
   d. AddRoundKey
4. Final Round (no MixColumns)
   a. SubBytes
   b. ShiftRows
   c. AddRoundKey

B.FOR DYNAMIC DATA SECURITY: RSA: COMMUNICATION SECURITY

The RSA algorithm is named after Ron Rivest, Adi Shamir and Len Adleman, who invented it in 1977[8].

Example: 1

1. Select primes p=11, q=3.
2. n = pq = 11.3 = 33 phi = (p-1)(q-1) = 10.2 = 20
3. Choose e=3
   Check gcd(e, p-1) = gcd(3, 10) = 1 (i.e. 3 and 10 have no common factors except 1),
   and check gcd(e, q-1) = gcd(3, 2) = 1
   therefore gcd(e, phi) = gcd(e, (p-1)(q-1)) = gcd(3, 20) = 1
4. Compute d such that ed ≡ 1 (mod phi)
   i.e. compute d = e^{-1} mod phi = 3^{-1} mod 20
   i.e. find a value for d such that phi divides (ed-1)
   i.e. find d such that 20 divides 3d-1.
   Simple testing (d = 1, 2 ...) gives d = 7
   Check: ed-1 = 3.7 - 1 = 20, which is divisible by phi.
5. Public key = (n, e) = (33, 3)
   Private key = (n, d) = (33, 7).

This is actually the smallest possible value for the modulus n for which the RSA algorithm works. Now say we want to encrypt the message m = 7,c = m^e mod n = 7^3 mod 33 = 343 mod 33 = 13.

RSA security is based on the difficulty of factoring large integers.

ISSN: 2278 – 1323 All Rights Reserved © 2015 IJARCET

C.ERROR PROCESSING: ERRORS ARE DETECTED DURING FAULT TREATMENT WHILE ACCESSING& STORING DATA.

MD5 CHECKSUM HASH PROPERTIES:

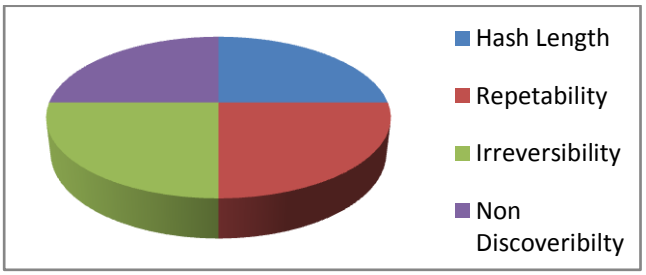

Fig 11: Showing Hash Properties

For integrity checking we require checksum. Computing for data in disk/data in cloud. Checksum are generated using hash function .MD5 hash algorithm is used to generate checksum, the MD5 Message-Digest Algorithm is a widely used cryptographic hash function that produces a128-bit (16-byte) hash value. A simple error-detection scheme in which each transmitted message is accompanied by a numerical value based on the number of set bits in the message. The receiving station then applies the same formula to the message and checks to make sure the accompanying numerical value is the same. If not, the receiver can assume that the message has been garbled. The algorithm takes as input a message of arbitrary length and produces as output a 128-bit "fingerprint" or "message digest" of the input. It is conjectured that it is computationally infeasible to produce two messages having the same message digest, or to produce any message having a given pre-specified target message digest. The MD5 algorithm is intended for digital signature applications, where a large file must be "compressed" in a secure manner before beingencrypted with a private (secret) key under a public-key cryptosystem such as RSA[9].

In java MessageDigest class is already present by which one can apply MD5 algorithm on data to obtained hash value of the data. so we have to create hashed value of this data need to create instance of MD5 by calling method "getInstance()" provide by that class MessageDigest," digest is used forobtaining hashed value of data[10].

Now compute checksum of hashed value obtained from MD5 algorithm:

```
For (int i=0; i<b.length;i++)
{
Result+=Integer.toString((b[i]&0x100,16).subst
ring (1)
}
Result result;
```

checksum obtained from above code saved in file, every time data is accessed & checksum is again created & match with saved checksum in this file. If it is matched then it is shows that data is not corrupted and can be used if it is not matched then it is shows that data is corrupted. Number of datacenters on cloud from datacenter 1 to n. If datacenter 1 is corrupted; data is downloaded from datacenter2, if no other datacenters are corrupted at that time data is return to user.

III. ARCHITECTURE AND IMPLEMENTATION

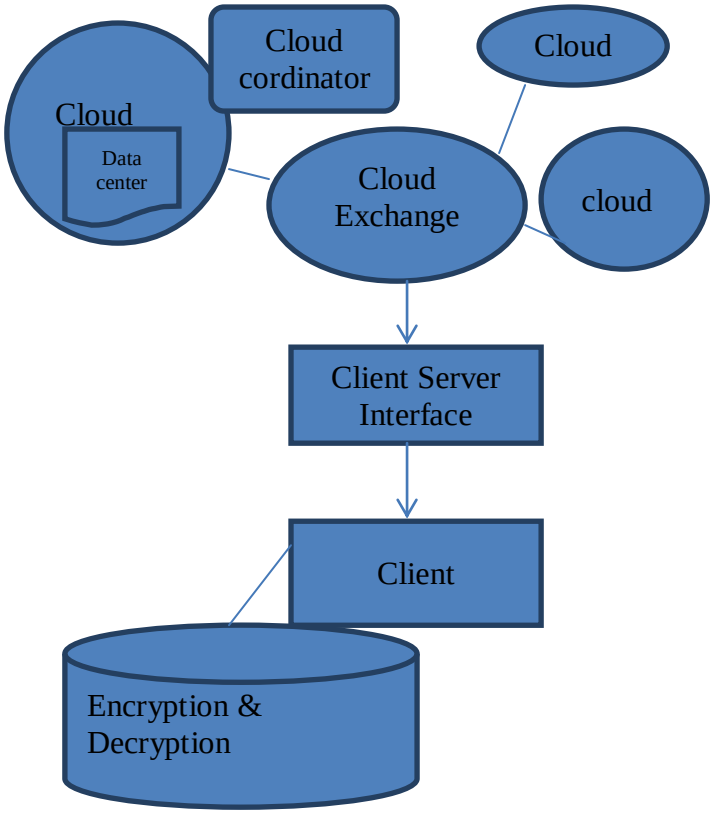

Fig 12: Architecture

As per as the Implementation is Concerned the Proposed model includes two basic and important parts that is modelling a cloud and secure cloud data storage.

On implementation include Cloud Simulation where we have implemented CryptoRSA class, CustoCrypto, UDPFileSend, UDPFileStore Class that associates with cloud exchange that include ccserver, cxserver, client thread, CloudInfo, U-Broker Class that associates with Cloud co-ordinator includes server,File/DataManager,CryptoRSA,DCFileStore,UDPFileSend,StorageManager Class, Checksum for checking every bit of data present on datacentre, DCFileSend, UDPFileStore.

ISSN: 2278 – 1323 All Rights Reserved © 2015 IJARCET

CloudSim	Crypto RSA CustomCrypto(AES) UDPFilStore
Cloud Exchange	cserver cxserver ClientThread UBroker CloudInfo
Cloud Coordinator	ccserver StorageManager FileManager Checksum CryptoRSA DCFileSend DCFileStore UDPFileSend UDPFileStore

VI. APPLICATION

Top Cloud Computing companies and Key Features:

Cloud Name Key Feature:

Sun Microsystems Sun Cloud	More available application than any other open os
IBM Dynamic Infrastructure	Integrated power management to help you plan, predict, monitor and activity manage power actively manage power consumption of your Blade Centre servers.
Amazon EC2	Designed to make wed-scale computing easier for developers
Google App Engine	No limit to free trial period if you do not exceed the quota allotted.
Microsoft Azure	Currently offering a "development accelerator" discount plan. 15-30 % discount off consumption changes for first 6 months.
AT&T Synaptic Hosting	Use fully on-demand infrastructure or combine it with dedicated components to meet specialized requirements.
Go Grid Cloud computing	Free load balancing and free 24/7 supports.
Sales force	Offers cloud solutions for automation and platform system performance and security at trust.salesforce.com

VII CONCLUSION AND FUTURE SCOPE

We conclude that the proposed method in above paper deals with security of Static data as well as dynamic data in cloud. AES is faster than RSA. RSA performance depends on prime number &complexity depends on that value of Primes thereforepaper gives us combining aspect of security, protection, detection of threads.

In future the same can be implemented for government organization , higher authority offices ,intelligent agency etc. With some other security algorithms depending on the requirement.

REFRENCES

[1]Sunil Sanka, ChittaranjanHota, MuttukrishnanRajarajan, "Secure Data Access in Cloud Computing," in IMSAA '10, 2010, p. 1-6.

[2]Chang Hsu and Chien-Hsing Wu, "A Business Model for Cloud Computing Based on a Separate Encryption and Decryption Service," in ICISA '11, 2011, p. 1-7.

[3] Chen, "Hiding Sensitive Association Rules with Limited Side Effects," in IEEE Transactions on Knowledge and Data engineering, Vol. 19, No. 1,pp. 29-42, January 2007.

[4]Sussman and Booz Allen alumnus " http://boozalen.com/2010".

[5]M.L.LilliBridge, S.Elnikety, A.Birrell, M.Burrows, M.Isard,"A Cooperative Internet Backup Scheme", *In Proceedings of the 2003 USENIX Annual Technical Conference, General Track*, pp.29-41, 2003.

[6]M.L.LilliBridge, S.Elnikety, A.Birrell, M.Burrows, M.Isard,"A Cooperative Internet Backup Scheme", *In Proceedings of the 2003 USENIX Annual Technical Conference, General Track*, pp.29-41, 2003.

[7]T.S.J.Schwarz and E.L.miller, "Store, Forgetm and check: Using Algebraic Signatures to Check Remotely Administered Storage", *In Proceeding of ICDCS'06*, pp.12-12, 2006.

[8]William Stallings "Cryptography and Network Security Principles and Practices", Prentice Hall, New Delhi.

[9]Popvickresimir, hocenski, zeljko,"Cloud Computing Security issue and challenges", *In Proceeding of 33rd International Convention, IEEE Transaction*, 2010.

[10]PriyankaArora, ArunSingh, "Evaluation and Comparison of Security Issues on Cloud Computing Environment", *World of Computer Science and Information Technology Journal, WCSIT*, ISSN: 2221-0741 Vol. 2, No. 5, 179-183, 2012

ISSN: 2278 – 1323 All Rights Reserved © 2015 IJARCET