Volume 4, Issue 3, March 2014 ISSN: 2277 128X

International Journal of Advanced Research in Computer Science and Software Engineering

Research Paper

Available online at: www.ijarcsse.com

Data Security in Cloud Computing

Prof Swarnalata Bollavarapu, Bharat Gupta

Department Of Computer Engineering, MPSTME

NMIMS University, India

Abstract— ***Cloud Computing has become a boon for an IT industry nowadays. It is like a next stage platform in the evolution of Internet. It provides a platform with an enhanced and efficient way to store data in the cloud i.e. server with different range of capabilities and application. It provides an easy way of accessing one's personal file or data and use application without installing it on machines by just having Internet access. We can have efficient computing by centralized data storage, processing and bandwidth. Example: Yahoo, Gmail, Amazon etc. are good cloud service providers. So all we need is to have Internet access then we can send mail and can access our account from any part of the world. The server and the email management software is installed on the cloud and managed by service providers. Providing an easy access to work and business still it have a major problem and threat i.e. "DATA SECURITY". Cloud has single layer security architecture and demand is high for customers. So this article mainly focuses on the study of algorithms used for data storage security in the cloud and desktops. And to overcome these problems encryption and decryption techniques like RSA and RC4 has been discussed here in more details.***

Keywords — ***Data Security, RSA, RC4, El-Gamal , Elliptic Curve and Digital Signature.***

I. INTRODUCTION

Cloud Computing is based on the concept of virtualization. In order words we can say that virtual computers are the components of cloud. The first question arises in ones mind is what you mean by cloud computing? Or what is cloud computing? People often confuse it with something in air or cloud that is cloud computing. But actually cloud computing is basic concept of separating everything like applications, software and even the infrastructure from the hardware you are working on. Ex. Google Doc is a classic web application, Google spreadsheet, Zen, Quick Books and many more. According to NIST definition of Cloud Computing "Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction."[1] The idea behind having cloud computing is if some how your system get crash or your windows got corrupted or some other fatal damage is done to your hardware then the software even the application on it get affected and one is left with nothing in hand. The one way is you can have entire backup of your system. But that is too costly not everyone can afford. As external hard disk are very expensive. The other way out is one can buy some storage from cloud service providers and can store their data. Next time if your system goes down or window is crashed then your data is not lost as it is secured up in the cloud i.e. server of the service providers. Nowadays major Cloud Service Providers are Amazon, Rackspace, Google, Microsoft, VMware, iCloud, Drop Box etc.

Cloud Computing Architecture

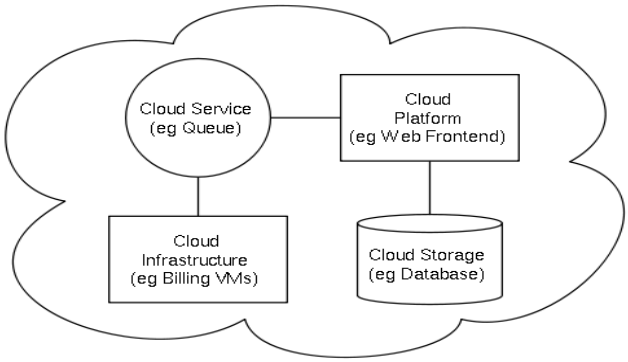

Fig 1: Basic Cloud Computing Architecture [2]

The above figure shows Cloud service provides service as a platform and service as an Infrastructure. Cloud Service deals with web-fronted applications by using various languages java, php etc. Cloud platform provides an environment for web-fronted applications. Cloud infrastructure provides user the remote infrastructures. The web fronted application are further connected to the database i.e. cloud storage.

ESSENTIAL CHARACTERISTICS [1]:

On-Demand Self Service: As the name says a consumer can demand for a service even without interacting with each service provider and can have computing capabilities like time and network storage as per the requirements.

Broad network access: Heterogeneous thin or thick client platforms like mobile phones, tablets, laptops and workstation access the capabilities available on the network through standard mechanism.

Resource pooling: In order to serve multiple consumers using a multi tenant model, with different physical and virtual resources given to them dynamically or taken back when work is done according to the consumer demand, the service providers pool their computing resources. Here consumer is not having any idea from where the resources are provided but able to specify location at higher level of abstraction (ex. country, state etc.). Some examples of resources that are pooled are storage, bandwidth, memory and processing.

Rapid Elasticity: Sometimes what happens that the consumers demand is very high at point of time, so in order to avoid delay at the consumer end the cloud has a most important feature called as scalability, the resources are provided to the consumer can be elastically provisioned and released.

Measured Service: Service provider charge users as per their usage in the server or charge-per-use basis. Types of services like storage, processing, bandwidth etc. are leveraged at some level of abstraction as cloud system automatically control and optimize resources. And the monitoring, controlling and reporting of resources usage are done in order to provide transparency for provider as well as for consumer using the service.

SERVICE MODEL

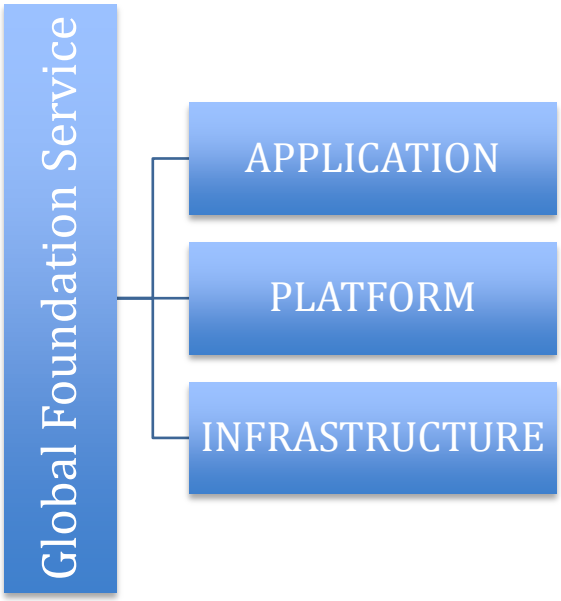

Fig 2: Service Model

Service model also called as SPI model as Software, Platform and Infrastructure Model. Software as a Service (SaaS): As the name says, it deals with the software or web based applications. Web based application are those applications that are built using web languages like php, java, .net, etc. this model of cloud allows one to run existing online applications. Ex. Google Docs. Platform as a Service (PaaS): Platform as a Service provides platform to users to work on web application or software. It allows users to create own cloud applications using supplier-specific tools and language. Ex Google App Engine Infrastructure as a Service (IaaS): users uses remote infrastructure, allows users to run any applications they want on cloud hardware of their own choice. Ex. Private cloud, dedicated hosting, hybrid hosting. Another Ex. Amazon provides elastic computing. One can ask for 1GB Storage, 256 RAM, 1 GB transfer/month server like this. Amazon EC2.

DEPLOYMENT MODEL

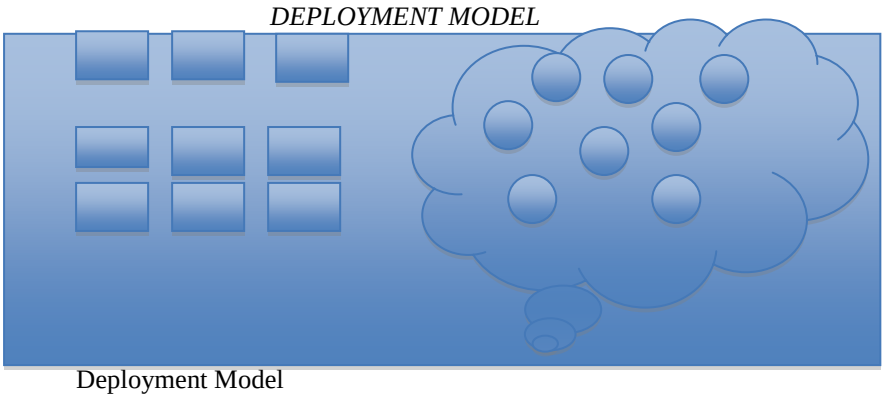

Fig3: Deployment Model

Note: Here, The block represents Physical Server and the circles Virtual Server.

Private Cloud: Having own physical server at a company or exclusive use by a single organization comprising multiple users.

Dedicated Hosting: It uses physical instances but not at a particular organization but can have many servers at different center in a whole campus or organization. Also called as Community Cloud.

Hybrid Hosting: When cloud is composition of two or more distinct cloud infrastructure thereby forming a unique entity but are bound by standardized that enables data portability, example cloud bursting for load balancing between clouds. It uses both physical and virtual instances.

Cloud Hosting: When cloud uses only virtual instances in the Vendor data Center.

II. VIRTUALIZATION

WHAT IS VIRTUALIZATION?

Virtualization technology enables to run multiple operating systems (or virtual machines) simultaneously on a single physical machine sharing the underlying resources. Some of the reason for using virtualization is: a) sufficient capability of recent computers to run multiple operating systems; b) using multiple isolated operating systems can maximize the use of resources; c) running different operating systems on single physical machine (for example Linux and Windows) [3] .It is a component of Cloud computing. It is like separating operating system from hardware or we can say that transfer of entire application to another hardware. Role of Virtualization in cloud computing Virtualization specifies creation of a virtual version of resources or advice, such as storage devices; a server, network or it can also be an operating system where the resource is divided by the framework into one or more execution environments. Something as simple as partitioning of a hard drive can be considered virtualization because the user takes one drive and partitions it to create two separate hard drives. Devices, applications and human users are able to interact with the virtual resource via virtualization as if it were a real single logical resource. Virtualization has two software i.e. Client Installed and Hypervisor. Client installed is the virtualized software and to manage this virtualized software we use V-sphere, this is used to administrate the computer having virtual software like ESxi or VMware fusion. Virtualization is very important aspect because for example if a consumer is in middle of work and system crash down and there is possibility of data been lost. So in order to avoid this virtualization comes into play. If data is stored in virtual servers, and then system crash down then there is no possibility of data been lost. E.g. Google spreadsheet

III.PROBLEMS IN CLOUD COMPUTING

Cloud Security: Basically deals with data in the cloud that might get affected by viruses or stolen by some other person. So the proper security is required for the same.

Cloud Manageability: Once the decision has been taken to use which type of cloud next step is to manage the deployment. There are many layers of cloud computing including cloud providers, development platform etc. So the companies have partnered with 'RightScale' for cloud management, which manages platform that delivers the scalable, cost-effective, on-demand power of cloud. Benefits of having Cloud manageability are avoids vendors lock in, supports both public and private cloud, add more flexibility, also provides automated scaling. Not only these benefits it also allows easy monitoring and management of environment.

Cloud Governance Compliance: In order to minimize the risk and at the same time keep moving towards their goal a organizational structure is required to keep an organization functional more in apt manner. Governance resolves the conflicts within an organization; this applies to the cloud providers as well as to the consumers.

DATA SECURITY ISSUES

- Data Integrity: As we know data integrity is very essential for data centers, similarly Integrity monitoring is vital in cloud servers. Corruption of data can take place anywhere in storage with any type of file. And the responsibility of managing the integrity of data falls on the company, which actually owns the data not the service providers.
- Data theft: Many copies are made of our data in the cloud. And also when one decides to move from one-service providers to others the problem of shifting entire data comes. So chances of data been stolen are more.
- Privacy Issues: These issues deals with the data privacy i.e. data should kept in a way that other illegal access to it wont effect the data. This can be done by encryption techniques.
- Infected Application: Sometimes the application we are working on is affected by viruses and that application we store in data centers of the providers that may effect other applications too. iCloud first scans all the data coming up the server for viruses and then store them in servers.
- Data loss: Data corruption can take place anywhere in the storage environment. While migrating of your data from one platform to other i.e. in the cloud. Cloud servers are big data centers that are having hardware and software, which are also prone to data corruption. Example: Amazon failure, many companies suffered downtime and also 0.07 percent actually lost their data.
- Security on user and vendor level: Security issues that resides at consumer levels are: Authentication, Data Protection, Loss of Governance whereas at vendor level the security issues are: Privacy, Securing Data in Transmission, Identity and access management.

IV. METHOD PROPOSED

Method Proposed in Paper [4]:

RSA:- Ron Rivest, Adi Shamir and Leonard Adleman described the RSA algorithm in 1978. The letter RSA is abbreviating form by initials of their surname. RSA algorithm involves three steps algorithm key generation, encryption and decryption. In this RSA algorithm, m is known as the modulus, „E" is known as the encryption exponent or public key exponent and „D" is known as the decryption exponent or private key exponent.

Algorithm [5]:

1. Choose two large prime P & Q
2. Calculate N = P * Q
3. Select the public key (i.e. encryption key) E such that it is not a factor of (P - 1) and (Q - 1).
4. Select the private key (i.e. decryption key) D such that following equation is true:
   (D * E) mod (P - 1) * (Q - 1) = 1
5. For encryption calculate cipher text CT from the plain text PT as fallows:
   $CT = PT^E \bmod N$
6. Send CT as the cipher text to the receiver.
7. For decryption, calculate the plain text PT from the cipher text CT as fallows:
   $PT = CT^D \bmod N$

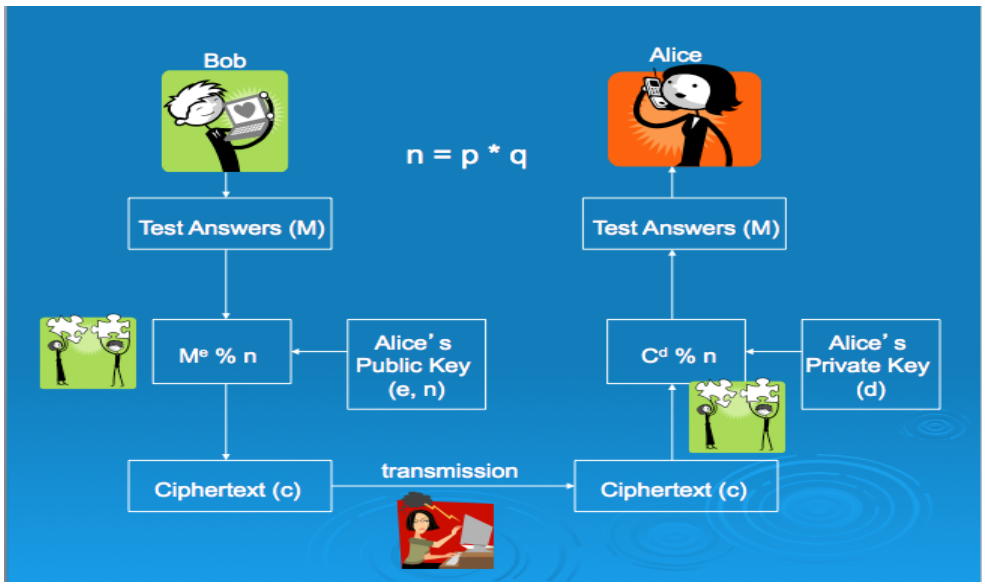

Fig 4: RSA Algorithm [6]

Method Proposed in Paper [7]:

RC4 is an encryption algorithm that was created by Ronald Rivest of RSA Security. It is used in WEP and WPA, which are encryption protocols commonly used on wireless routers. The workings of RC4 used to be a secret, but its code was leaked onto the internet in 1994. RC4 was originally very widely used due to its simplicity and speed. Typically 16 byte keys are used for strong encryption, but shorter key lengths are also widely used due to export restrictions. Over time this code was shown to produce biased outputs towards certain sequences, mostly in first few bytes of the keystream generated. This led to a future version of the RC4 code that is more widely used today, called RC4-drop[n], in which the first n bytes of the keystream are dropped in order to get rid of this biased output. Some notable uses of RC4 are implemented in Microsoft Excel, Adobe's Acrobat 2.0 (1994), and BitTorrent clients. [8]

Algorithm [7]

A. The Key-Scheduling Algorithm (KSA)

The key-scheduling algorithm is used to initialize the permutation in the array "S". "Key Length" is defined as the number of bytes in the key and can be in the range 1 ≤ key length ≤ 256, typically between 5 and 16, corresponding to a key length of 40 – 128 bits. First, the array "S" is initialized to the identity permutation. S is then processed for 256 iterations in a similar way to the main PRGA algorithm, but also mixes in bytes of the key at the same time. First stage of KSA algorithm steps as explained as following.

1. Initialize the variable length key "i" of size from 0 to 255.
2. Initialize the key matrix "S[i]" as per the size of the input key.
3. Initialize the state table of fixed size 256 bytes from the value 0 to 255 in ascending order.
4. Using the key matrix of variable size done by the permutation on the "S" table
5. Output of the KSA, the final "S" table after swap operation.

KSA algorithm

for i from 0 to 255 S[i] := i
endfor j := 0 for i from 0 to 255
j := (j + S[i] + key[i mod key length]) mod 256 swap (&S[i],&S[j]) endfor

B. Pseudo-Random Generation Algorithm (PRGA).
int i=0, j=0 forjfrom0 to 255) for i from 0 to 255) i=(i+1) mod 256 end for
j=(j + s[i]) mod 256 swap (s[i], s[j]) end for output= s[s[i] + s[j]] mod 256]

Method Proposed in Paper [9 & 10]:
Neal Koblitz [11] and Victor Miller [12] as applied to cryptography first proposed elliptic Curve (EC) systems in 1985 independently. ECC is a public-key cryptosystem. Elliptic curves are used as an extension to other current cryptosystems. That is Elliptic Curve Diffie-Hellman Key Exchange and Elliptic Curve Digital Signature Algorithm. ECC depends upon the hardness of the discrete log problem.

E: $y^2 \square x^3 \square ax \square b$ (equation for Elliptic Curve)
If P1 and P2 are on E,
P3 = P1 + P2

As shown in picture. Let P1=(x1, y1), P2=(x2, y2), P3=(x3, y3) and P1 not equals P2

$$m = \frac{y_2 - y_1}{x_2 - x_1};$$

To find the intersection with E. we get

$$(m(x - x_1) + y_1)^2 = x^3 + Ax + B$$

$$or, 0 = x^3 - m^2x^2 + \ldots$$

$$So, x_3 = m^2 - x_1 - x_2$$

$$\Rightarrow y_3 = m(x_1 - x_2) - y_1$$

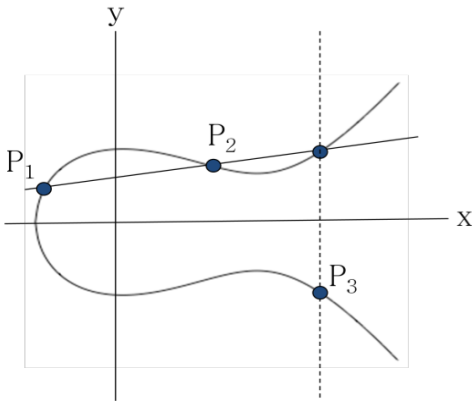

Fig 5: Elliptic Curve [10]

ALGORITHM FOR DATA SECURITY USING ECC

- Both clouds agree to some publicly known data item.

1. The elliptic curve equation
   - values of a and b
   - prime, p
2. The elliptic group computed from the elliptic curve equation
3. A base point, B, taken from the elliptic group

Key generation:
6. A selects an integer dA. this is A's private key.
7. A then generates a public key PA=dA*B
8. B similarly selects a private key dB and computes a public key PB= dB *B
9. A generates the security key K= dA *PB. B generates the secrete key K= dB *PA.

Signature Generation:
For signing a message m by sender of cloud A, using A's private key dA
1. Calculate e=HASH (m), where HASH is a cryptographic hash function, such as SHA-1
2. Select a random integer k from [1, n − 1]
3. Calculater=x1(modn),where(x1,y1)=k*B.Ifr=0, go to step 2
4. Calculates=k−1(e+dAr)(modn).Ifs=0,gotostep2
5. The signature is the pair (r, s)
6. Send signature (r, s) to B cloud.

Encryption algorithm:
Suppose A wants to send to B an encrypted message.
i. A takes plaintext message M, and encodes it onto a point, PM, from the elliptic group.
ii. A chooses another random integer, k from the interval [1, p-1]
iii. The cipher text is a pair of points PC = [(kB), (PM + kPB)]
iv. Send ciphertext PC to cloud B.

Decryption algorithm:
Cloud B will take the following steps to decrypt cipher text PC.
a. B computes the product of the first point from PC and his private key, dB dB * (kB)
b. B then takes this product and subtracts it from the second point from PC
(PM + kPB) – [dB(kB)] = PM + k(dBB) – dB(kB) = PM
c. B cloud then decodes PM to get the message, M.

Signature Verification:

For B to authenticate A's signature, B must have A's public key PA

1. Verify that r and s are integers in [1, n − 1]. If not, the signature is invalid
2. Calculate e = HASH (m), where HASH is the same function used in the signature generation
3. Calculate w = s −1 (mod n)
4. Calculate u1 = ew (mod n) and u2 = rw (mod n)
5. Calcúlate (x1, y1) = u1B + u2PA
6. The signature is valid if x1 = r(mod n), invalid otherwise.

ElGamal Algorithm:

And also ElGamal technique [5] is used for both:

- Digital Signature
- Encryption

To generate a key pair:

1. Select a prime no. p
2. Select two random nos. g & x (such that g & x < p)
3. Find: $y = g^x$ mod p (public keys are y, g, p and private key is x)
4. For encrypting a plaintext message M, select random no. K (such that K is relatively prime to (p-1))
5. Then find:
   - $a = g^k$ mod p

   M = (ax + kb) mod (p - 1)

   Note: Find 'b' using above equation
6. Cipher Text = (a, b) (double the size of plaintext)
7. Decryption: calculate M
   - $M = b/a^x$ mod p

     where: M = plaintext

V. INFERENCES

TABLE 1

S.No	Asymmetric Key/Public Key Cryptography	Symmetric Key
1.	Use different key for Encryption & Decryption. Here public key and private key are used.	Use same key for Encryption & Decryption. Here secret key is used.
2.	Speed of Encryption & Decryption is slow	Speed of Encryption & Decryption is fast
3.	Size of resulting encrypted text more than the original clear text size	Size of resulting encrypted text usually same as or less than the original clear text size.
4.	Key exchange: No problem	Key exchange: Big problem
5.	Usage: Encryption & Decryption and Digital Signature	Usage: Encryption & Decryption only
6.	EX: RSA, ECC, ElGamal	EX: RC4

- Advantages of RSA:

1. RSA's biggest advantage is that it uses Public Key encryption. This means that your text will be encrypted with someone's Public Key.
2. Increased security and convenience.
3. Provides digital signatures that cannot be repudiated.

- Disadvantages of RSA:

1. Slower than secret key method, but can be used in conjunction with the secret key to make it more efficient.
2. Can be vulnerable to impersonation if hacked.

- Advantages of RC4:

1. Not patent i.e. authority to use for public application development.
2. Security provided.
3. Possibility of implementation on selected device.
4. Variable key length algorithm.
5. Easy for hardware development

- Disadvantages of RC4:

1. Vulnerable when the beginning of the output key stream is not discarded, or when nonrandom or related keys are used.

- Applications -RC4-based cryptosystems:

1) WEP
2) WPA
3) Cipher Saber
4) Bit Torrent protocol encryption
5) Microsoft Point-to-Point Encryption

- Advantages of ECC:

1. Greater flexibility in choosing cryptographic system
2. No known sub exponential time algorithm for ECDLP. Smaller key sizes (with the same security). The minimum key size for ECC should be 132 bits vs. 952 bits for RSA. Shorter key length translate into faster handshaking protocol.
3. As a result: greater speed, less storage ⇒ECC can be used in smart cards, cellular phones, pagers etc.
4. ECC over RSA are particularly important in wireless devices, where computing power, memory and battery life are limited.

- Disadvantages Of ECC:

1. Hyper elliptic cryptosystems offer even smaller key sizes.
2. ECC increases the size of the encrypted message significantly more than RSA encryption.
3. ECC is mathematically more subtle than RSA
4. Difficult to explain/justify to the client, more complex thereby reducing the security of the algorithm

- Applications of ECC:

1. Key exchange, digital signature, authentication, (limited) message transmission, etc.
2. Used in Smart Cards without mathematical processor.
3. Contactless Smart Card works with ECC because other software requires more energy.

- Advantages Of ECC:

1. El-Gamal has homomorphic property that makes it useful in application like e-voting

- Disadvantages Of El-Gamal:

1. Need for randomness
2. Slow speed, especially for signing.
3. Message expansion by the factor of two takes place during encryption.

VI. CONCLUSION

Cloud Computing is the phenomenon of separating applications from hardware and providing an easy way to deploy on demanded server. Service models: Application, Platform and Infrastructure. The functioning of Cloud Computing is greatly affected by issues such as that of data security, integrity, theft, loss and presence of infected applications. To solve these issues various algorithms such as ECC, RSA, RC4 and El-Gamal have been suggested and discussed in the paper so that any unauthorized user trying to access the confidential data will not be allowed to access the cloud. ECC (Elliptic Curve Cryptography) involves Deffie-Hellman Key Exchange and Elliptic Curve Digital Signature Algorithm. It is a public key algorithm, which performs both digital signature and encryption. Here, Security is based on the difficulty of computing discrete logarithm in a finite field. El-Gamal, ECC and RSA are forms of public key cryptography, in which one encryption key, known as the private key, is kept secret, while another, known as a public key, is freely distributed. Public key cryptography is computationally more expensive than private key encryption, which employs a single, shared encryption key. In wireless devices, public key encryption can shorten the lifetime of batteries or of the devices themselves. Problem with public key exchange is “MAN IN MIDDLE ATTACK”.

REFERENCES

[1] NIST National Institute Of Standards and Technology U.S. Department Of Commerce.
[2] http://www.slideshare.net/ronak2454/issues-in-cloud-computing-9710875
[3] Krishnadhan Das “VirtualCloud - A Cloud Environment Simulator” Department of Computer Science and Engineering Indian Institute of Technology, Bombay
[4] Nilesh N. Kumbhar , Virendrasingh V. Chaudhari , Mohit A.Badhe , “The Comprehensive Approach for Data Security in Cloud Computing: A Survey”, International Journal of Computer Applications (0975 – 8887) Volume 39– No.18, February 2012.
[5] Atul Kahate “Cryptography and Network Security” Tata McGraw-Hill.
[6] Kiera Caponi, Larissa Grayson “RSA Cryptography” http://www.codeproject.com/dotnet/RSACryptoPad.asp, http://en.wikipedia.org/wiki/Rsa, Cormen, Leiserson, Rivest, Stein. Introduction to Algorithms. McGraw-Hill.
[7] Vijay. G.R, A.Rama Mohan Reddy “ Data Security in Cloud based on Trusted Computing Environment”, International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-3, Issue-1, March 2013.
[8] Ralph (Eddie) Rise Suk-Hyun Cho Devin Kaylor “RC4 Encryption.pdf”

[9]Fuwen Liu "A Tutorial on Elliptic Curve Cryptography (ECC)" lfw@informatik.tu-cottbus.de -Brandenburg Technical University of Cottbus

[10] Veerraju Gampala, Srilakshmi Inuganti, Satish Muppidi "Data Security in Cloud Computing with Elliptic Curve Cryptography" International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-2, Issue-3, July 2012.

[11] N. Koblitz. Elliptic curve cryptosystems. Mathematics of Computation, 48:203–209,1987.

[12] V. Miller. Use of elliptic curves in cryptography. Advances in Cryptology—CRYPTO '85 (LNCS 218) [483], 417– 426, 1986.

