RESEARCH ARTICLE OPEN ACCESS

A Survey On Various Algorithms For Data Security In Cloud Computing Environment

Savita.A.Harkude [1], Dr. G.N.Kodanda Ramaiah [2]
Department of T.C.E [1]
SirMVIT, Bangalore
Department of E.C.E [2]
Kuppam Engineering College, Kuppam
India

ABSTRACT
Data security is a major issue in cloud computing now a day as data is stored in different locations. Here we make a comparative research on the existing work regarding the data security and privacy protection techniques used in the cloud computing. There are two main user concern issues regarding data security and privacy protection which are related to hardware and software in cloud computing. Storing of data onto cloud in a secure and safe manner in order to avoid intrusions and attacks which reduce the cost and time to store the encrypted data in Cloud computing has been facing lots of security issues regarding privacy of data. Data mining is considered to be essential component in business domain.
Keywords:- Cloud computing, Data Security, Privacy Protection.

I. INTRODUCTION

Cloud computing is an architecture that provides computing service through the internet on demand and it also provides user access pool of shared resources namely networks, storage, servers, services and applications. Cloud computing allows the user with storing and sharing of data in large amount. Cloud provides the user high speed data transfer services through the internet. Cloud also provides additional features to users like collect data or share data from anywhere through internet. Cloud computing allows multiple users to access single server to perform different operations on their data. Cloud also provides high speed services at minimum cost. Users can access and store the data in cloud from remotely. Store data in the cloud so that data is easily available to the customers and deliver the data efficiently to customers. Service provider provides service to the customer who can easily recover data that has been hacked or lost. And also by using encryption technique, it provides data authentication so that user does not face any issues like data loss or data theft.

A. Service models in cloud computing.

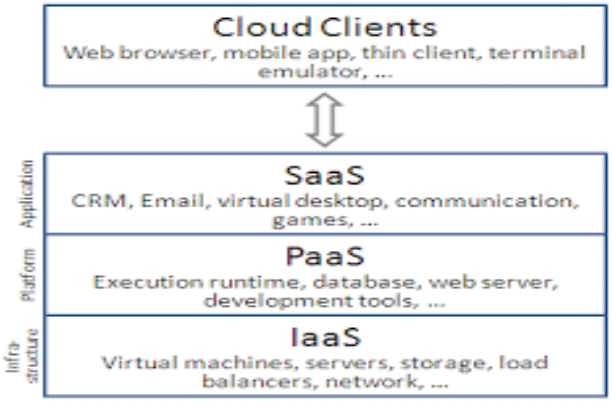
Software as a Service (SaaS): It access service in cloud computing by using software like browser like gmail and other services from google.

Platform as a Service (Paas): It allows the user to access service to develop and deploy the new applications.
Infrastructure as a Service (IaaS): Here user can access the service of servers computational and storage in cloud.

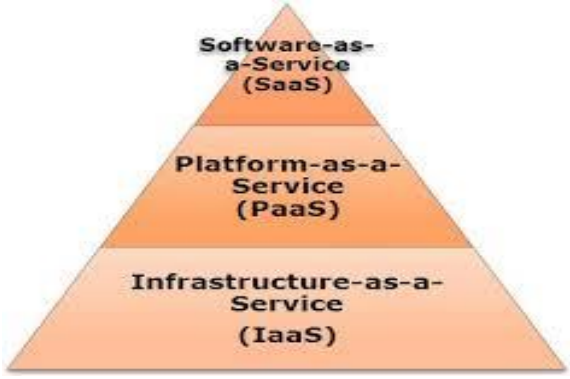
The remaining paper is as follows: section II presents literature survey. Section III describes Security issues and threats in cloud computing. Section IV deals with various algorithm approaches to provide security of data. Section V describes related work. Section VI describes conclusion.

II. LITERATURE SURVEY

Sumithra R [1] has surveyed the security problems on



cloud computing and outsourcing data mining. Kiril alexiev [2] discuss some results from the cryptanalysis of



the IDA encryption algorithm. He compares with the most prominent and well behaved encryption algorithms. Rohit bhaduria [3] surveys elaborate and analyze the numerous unresolved issues threatening the cloud computing and affecting the various stake holders associated with it. Anjali V. Almale [4] surveys data mining in cloud computing. Some key features are used for distribution of data in some ways that is understood by user. It proposes to implement cloud security aspects for data mining in cloud system. Kajal Rani [5] proposed the scheme of cloud data storage security strategy to overcome the problems of traditional data protection algorithm and improving security of data in cloud computing using steganography, encryption decryption techniques, compression and splitting technique. Katarzyna kapusta [6] proposed fast fragmentation algorithm for data protection in multi cloud environment. They performed an empirical security analysis on data sets provided by a large enterprise and results shows good protection. This scheme also performs twice faster than previous fragmentation techniques. Elham Abd [7] surveyed on enhancement of data security in cloud computing by using data segregation. They show the comparison between different segregation techniques pros and cons. Subbiah S. Selva Muthukumaran [8] proposed security strategy in cloud data storage using vertical partitioning algorithm. This algorithm provides data security in efficient manner. Zebin et al.,[9] implements PCA algorithm on SPARK and discusses risk of exploiting cloud architecture for distributed and parallel dimensionality reduction and he used data repositories for storing remotely sensed datasets. Bermer [10] combines cloud and

peer to peer computing that contains backup, online gaming, content distribution and streaming.

III. SECURITY ISSUES IN CLOUD - COMPUTING

A. ***In cloud environment, there are different security issues as discussed below.***

Fig 1 Data security issues and threats

Data Protection:
The data of each customer must be properly kept apart from other customer and it must be able move to securely from one location to another. Cloud computing prevents data loss and access by others.
Authentication:
Avoid IP spoofing by using encrypted protocols wherever possible. Avoid ARP poisoning by using root access to change ARP tables.
Data Verification:
There is a chance of tampering, theft or loss of data by third parties on local machine or in transit or at rest and during backups on remote. Resource isolation ensures security of data during processing, by isolating the processor caches in virtual machines, and isolating those virtual caches from the Hypervisor cache.

Infected Application:
Uploading of infected application by malicious user into cloud affects the customer data. Here it requires application security measures like firewall in production environment.
Availability:
Cloud providers ensure that data and application are regularly available to customer.

B. Data Security in Cloud Environment must achieve the following objectives:

Integrity of Data: data integrity means guarantee that data is uncorrupted and only authorized user can access or modify the data. More accuracy, consistency and trustworthiness are required to maintain data over life cycle. Data should not be changed during transit and some important steps should be taken to prevent data alteration from unauthorized user.

Data Confidentiality: Data confidentiality is secure sensitive data within the cloud and also protects data from unintentional or unauthorized access. It is often a measure of the ability of system to protect data.

Data Availability & Management: It is ensuring that available of data to users and application where and whenever data is needed. Essential IT and management procedures, tools and technologies are needed to enable,

manage to make data available. Service level agreements are created and third party service provider provides assurance of service for available of data. Storage area networks (SAN) and RAID-based storage systems are some famous storage management technologies for availability of data.

Data Authenticity: Data can be assumed to authentic only when it has not been corrupted and also means that the data received from server is original and exactly.

Data Storage & Maintenance: Cloud storage is a model of computer data storage where the data is stored in logical pools. The storage may be multiple servers that may be placed in multiple locations. The responsibility of cloud storage provider is available and access of data. User can buy or lease storage capacity from provider for data application.

IV. DIFFERENT ALGORITHM APPROACHES

GZIPSTREAM AND DEFLATE ALGORITHM: It uses combination of LZ77, Huffman codes and run-length-encoding. The LZ77 algorithm replaces recurring occurrence of data with references. Each reference has two values one is the jump and the length.

Example: for text

```
Original text: "ServerGrove, the PHP hosting company, provides hosting
solutions for PHP projects" (81 bytes)
LZ77: "ServerGrove, the PHP hosting company, p<3,32>ides<9,26>solutions
for<5,52><3,35>jects" (73 bytes, assuming that each reference is 3 bytes)
```

The words hosting and PHP are second times occurred. That ways they are replaced by a reference. Similarly “er” sub string is repeated. Now original text remaining is

```
Original text: "ServerGrove"
ASCII codification: "01010011 01100101 01110010 01110110 01100101 01110010
01000111 01110010 01101111 01110110 01100101" (88 bits)
```

Huffman coding is variable-length coding technique that allocates short codes for repeated characters. The main issue of this coding is we must know the end of the code and when the new person for decoding. This is solved by adding Prefix code and no codeword is a prefix of another word.

```
Huffman: "1110 00 01 10 00 01 1111 01 110 10 00" (27 bits)
```

ASCII is a fixed-length character encoding method, letter e occurs three times in above text, but letter “G” occurs only one time. Huffman can create a most optimized scheme.

The Huffman technique permit us to get smaller codes for “e”, “r” and “v”, while “S” and “G” got the longer ones.

IDA ENCRYPTION ALGORITHM:

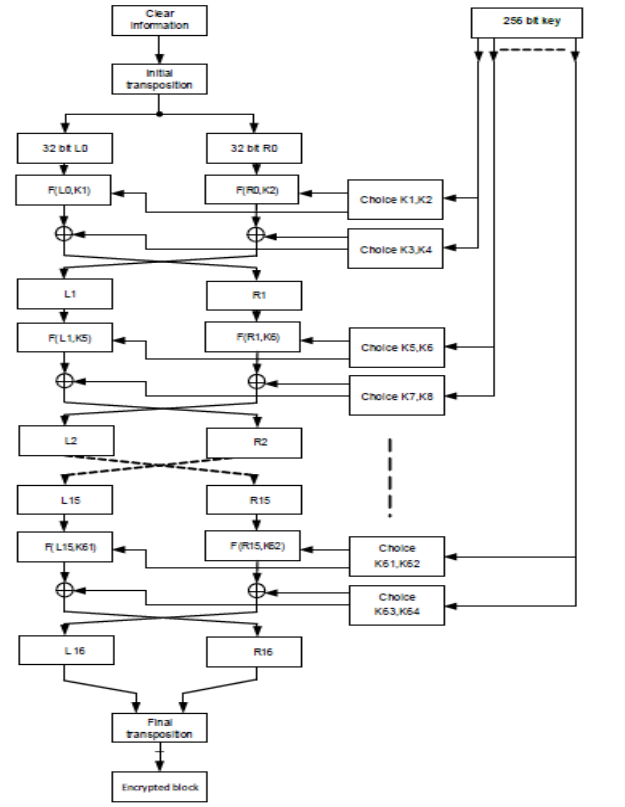
IDA is a 64-bit symmetric block cryptographic algorithm using 256-bit cryptographic key. It consists of 16 cycles containing transpositions, substitutions and nonlinear procedures. This algorithm involves encryption and decryption of 64-bits data blocks using 256-bit key and contains bitwise, logical operations and table permutations. [9]

58	50	42	34	26	18	10	2
60	52	44	36	28	20	12	4
62	54	46	38	30	22	14	6
64	56	48	40	32	24	16	8
57	49	41	33	25	17	9	1
59	51	43	35	27	19	11	3
61	53	45	37	29	21	13	5
63	55	47	39	31	23	15	7

40	8	48	16	56	24	64	32
39	7	47	15	55	23	63	31
38	6	46	14	54	22	62	30
37	5	45	13	53	21	61	29
36	4	44	12	52	20	60	28
35	3	43	11	51	19	59	27
34	2	42	10	50	18	58	26
33	1	41	9	49	17	57	25

The two sequences left L (0) and right R (0) which contains 32bits constructed from obtained bit sequence. By using function F (.), 32-bit sequences are encrypted. Encryption method contains two parts. First, the outputs of both F (.) 32 bit are fed to inputs of 2 XOR adders. Then second inputs of adders are connected with 32-bit sub keys. The outputs of two XOR adders change their places i.e. the right one becomes the left one and vice versa. This is the input sequence for next cycle. This is repeated for 16 cycles in order to provide secure data transfer. The flow diagram of IDA encryption is shown below. [2]

Generation of the 64 bit keys {Kj}



FAST FRAGMENTATION ALGORITHM:

Data processing flow:

Data d initially treated as combination of l data chunks. The l data chunks are encoded into l data share through one

by one. Then from data shares n fragments are constructed in such way that k fragments are sufficient to recover the data. For efficient processing, data chunks are reorganized into Data Chunk Sets of k elements, where DCSi(j) is the jth data chunk in set i. The data consists of DCSs: DCS1…….,DCSm (m=[1/k]). As a result encoding consists of DSSs: DSS1, DSSm. finally data shares spread to k final fragments and redundant fragments (n-k) are added.

$DCS_1(1)$	$DCS_1(2)$	$DCS_1(3)$
DCS2(1)	$DCS_2(2)$	$DCS_2(3)$
DCS3(1)	$DCS_3(2)$	$DCS_3(3)$

DSS0(1)	DSS0(2)	DSS0(3)
DSS1(1)	DSS1(2)	DSS1(3)
DSS2(1)	DSS2(2)	DSS2(3)
DSS3(1)	DSS3(2)	DSS3(3)

Example for k = 3: data chunks are transformed into data shares. Encoding of DCSi is based on DSSi−1. A pseudorandom seed serves as DSS0. DSS1 (2) comes from a transformation of DCS1 (2). DSS0 (1) and DSS0 (3) were used as coefficients of the encoding polynomial

Algorithm Fragmentation procedure
1: d = DCS1, ...,DCSm
2: DSS0 = s1, ..., sk
3: for i = 1 : m do
4: for j = 1 : k do
5: x = j _ DSSi−1(j)
6: if x == 0 then x = 1
7: end if
8: Coeffs = DSSi−1 \ DSSi−1(j)
9: DSSi(j) = Encode(DCSi(j),Coeffs, x)
10: end for
11: end for
12: f1, ..., fk = Distribute Shares(DSS0, ...,DSSm)
13: fk+1,..., fn = Add Redundancy(f1, ..., fk)

V. RELATED WORK

Katarzyna et al proposed dispersal algorithms to secure the data by diving data d into n fragments of size dsize/k where k is needed for data recovery. Data is initially processed by set of k chunks. Encoding the data is not based on multiplication of matrix, but on the modification of Shamir's secret sharing scheme. They used pseudo-random as first set of data chunks and distributes within data.

Novel fragmentation algorithm combines the keyless property, computational simplicity and space efficient size of fragment of information dispersal algorithms. Their work shows that produced fragments achieve good randomness and correlated neither with initial data nor among themselves. Fresh seed is used for fragmentation which ensures that backward and forward secrecy properties. This scheme can be applied to all application of data storage or transmission.

Kiril et al proposed some results of cryptanalysis IDA (Ivanov,divok,arnadov) algorithm. Their work is to compare all algorithms with IDA by using advanced cryptanalysis tools. IDA is 64-bit systemic block cryptographic algorithm using 256-bit cryptographic key which consists of 16 cycles containing transpositions, substitutions and non linear procedures.

Their work demonstrated its correspondence to contemporary state of art encryption algorithms. IDA is inherited best from old encryption algorithm like DES. To prevent brute force attack, good properties a longer key added and some additional are required to assure in the implementation in noisy environment.

Kajal et al proposed a scheme to provide the security for cloud data by using steganography, encryption decryption techniques, compression and splitting technique. They used steganography for concealment of data, messages etc with in computer. The data is splitting different parts with some extensions and then encrypt the data using encryption techniques DES.AES, RSA to prevent unauthorized user or hackers. GZIPSTREAM algorithm is used to compresses split files to reduce the size upload into cloud .zip based on the Deflate algorithm.

Comparison table between three algorithms

Algorithm depends upon the key management, type of cryptography, number of keys. Number of bits used in a key Longer, the key length and data length more will be the power consumption that will lead to more heat dissipation. From this examination and study we have shortlisted Fast Fragmentation. This algorithm is more secure and fast encryption.

PARAMET-ER	GZIPSTREA-M AND DEFLATE	IDA	FAST FRAGMEN T-ATION
Key length	Variable	64 bits	128 bits
Blocks	Variable	32	Depends on no of fragments
Attacks found	Breach attack	Brute force key attack	Brute force
Security level	Secure	Secure	Highly secure
Encryption speed	Slow	Normal	Fast
Application	In all https application	In integrated circuits	Multi-cloud environment

VI. CONCLUSION

The major challenges in cloud computing are security issues and threats. Many researchers are proposed various algorithms for data protection and to enhance data security. So far we surveyed various algorithms such IDA, fragmentation and gzipstream & deflate algorithms. More works are needed to secure data in cloud.

REFERENCES

[1]. R. Sumithra & Sujni Paul " A SURVEY PAPER ON CLOUD COMPUTING SECURITY AND UTSOURCING DATA MINING IN CLOUD PLATFORM" International Journal of Knowledge Management & e-Learning Volume 3 January-June 2011.

[2]. Kiril Alexiev1,2, Ivan Ivanov1, Krasimira Ivanova1, Emilia Saranova1 "Cryptanalysis of IDA encryption algorithm"

[3]. Rohit Bhadauria and Sugata Sanyal "Survey on Security Issues in Cloud Computing and Associated Mitigation Techniques"

[4]. Anjali V. Almale1, S.V. Phulari2, Shwets Shanwad "A Survey of Data Mining in Cloud Computing" DOI 0.4010/2016.1136 ISSN 2321 3361 © 2016 IJESC

[5]. Kajal Rani1, Raj Kumar Sagar2 "Enhanced Data Storage Security in Cloud Environment using Encryption, Compression and Splitting technique" 2017 2nd International Conference on Telecommunication and Networks (TEL-NET 2017)

[6]. Katarzyna Kapusta and Gerard Memmi A Fast Fragmentation Algorithm For Data Protection In a Multi-Cloud Environment arXiv:1804.01886v1 [cs.CR] 5 Apr 2018

[7]. Elham Abd Al Latif Al Badawi1 & Ahmed Kayed2 SURVEY ON ENHANCING THE DATA SECURITY OF THE CLOUD COMPUTING ENVIRONMENT BY USING DATA SEGREGATION TECHNIQUE IJRRAS 23 (2) May 2015

[8]. S. S. Muthukumaran and T. Ramkumar "An Approach for Enhancing Secure Cloud Storage Using Vertical Partitioning Algorithm", Middle-East Journal of Scientific Research, vol.23,no. 2, pp. 223-230, 2015.

[9]. Ivan Ivanov,Stella Vetova, Georgi Stanchev "Cryptographic Algorithm IDA for Security and Data Storage Increase in the Integrated Application of the European System " Ecall" For Automatic Emergency Calls in Car Accidents" IJEIT Volume 4,Issue 12,June 2015.

[10]. Bremer and K. Graffi, "Symbiotic Coupling of P2P and Cloud Systems: The Wikipedia Case", In the Proceedings of IEEE International conference on communication, pp. 3444-3449, 2013.

[11]. Vidyanand Ukey, Nitin Mishra " Dataset Segmentation for Cloud Computing and Securing Data Using ECC" Vidyanand Ukey et al, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (3) , 2014, 4210-4213

[12]. Sushmita Ruj‡, Milos Stojmenovic†, Amiya Nayak∗ "Decentralized Access Control with Anonymous Authentication of Data Stored in Clouds" IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS VOL:25 NO:2 YEAR 2014