JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

# Security of Big Data in Cloud

**Kulbir Kaur Sandhu[1], Navdeep Singh Gill[2]**

Assistant Professor[1], Scholar[2]
Department of Computer Science[1], Masters of business administration in maritime logistics and management[2]Baba Farid College, Bathinda[1], University of Tasmania[2]

Kulbirkaursandhu01@gmail.com[1], nsgill1987@gmail.com[2]

***Abstract*—Cloud computing requires various services to tackle the safety throughout the transmission of sensitive data and crucial applications to share and public cloud environments. The cloud environments are scaling massive for knowledge processing and storage. Cloud computing environments have various advantages as well as negative aspects on the data safety provider. In cloud computing networks, security is the main step to be improved to enhance performance of the system. In this paper, we have analyzed various challenges for the security of big data in cloud. Further, their suitable solutions are also discussed. It is assumed that ABE is the best scheme to enhance security of the cloud.**

***Keywords*—Encryption, Asymmetric Encryption Standard, Attribute based Scheme, Key Policy, Cipher Text, Trust Management**

## I. INTRODUCTION

With the advent of technology, there is a huge and explosive demand in computing and applications models. This demand helps in the growth of cloud services such as software as a service, network of community and web store. In today's technology era, cloud computing has become significant research topic. This latest technology provides internet based cluster system with wide range of services. Furthermore, Cloud computing is responsible for the revolution in IT industry by providing flexibility, pay for only used resources and services using by them. Services of cloud vary significantly in particular technology and its implementation. The best services of cloud are cost reduction, elasticity, pay-as-you-go and improved availability.

From last decades, a new way has been opened for storage and information in the form of cluster and grid computing. Through this, it is possible to store data on cloud with the help of loosely coupled grid.

Therefore, a new concept of cloud computing has been generated from clusters. This concept has somefeatures of networking but it provides on-demand services to its users. No doubt, cloud computing provides enormous services like reliability, flexibility and low cost services but still security of big data in cloud is a big issue. Security issues include risk of malicious insider, low trust level, privacy, confidentiality and encryption. Big data sizes are continually growing, currently ranging from a number of dozen terabytes (TB) to many petabytes (PB) of information in a single information set. With the advancement of technology and use of more data in companies there is a need faster and efficient approach to analyze this data [3]. Big knowledge Analytics (BDA) may additionally allow the construction of predictive models for client conduct and purchase patterns, results in raising overall profitability [4]. Enormous information can create transparency, and make primary information moreeasily accessible to stakeholders with time frame.In next section, we will discuss about various security issues for big data in cloud followed by the solution.

## II. SECURITY ISSUES OF BIG DATA IN CLOUD

In this section various security issues related to Big Data in Cloud has been discussed. These issues are:

### *A. Trust*

Nowadays, the main issue which is faced by the customer and service provider in cloud computing is Trust. There is always doubt in the mind of the user whether his/her data is secure from attacker or not and whether services are trustworthy or not. SLA bounded customer and services provider together by mention ,.m l vy,defined for SLA. Therefore, many efforts have been done in this field to resolve security and privacy issues. A trust model is discussed in [5] for the enhancement of the security of the cloud. A trust based [6] rating mechanism is presented to secure collaboration of cloud with social media.

### *B. Integrity*

After Trust another important issue is integrity. Improper modification of information comes under this issue [6]. Data resides in number of places in a cloud, therefore, access

control mechanism should be much protected and each user must be confirmed as an authentic user. Digital Signature is solution of this problem.

### C. *Confidentiality*

Confidentiality means to avoid the release of secretive and important information. Although data is stored at different geographical locations, therefore, confidentiality becomes big issue [7]. There is variety of methods like homomorphic encryption, distributive storage, data concealmentto secure confidentiality of data [21]. But it is an expensive method.

### D. *Encryption*

To secure the data in cloud computing, encryption is the most secure method. But it needs high level of computations. Every time data is decrypted before processing, therefore, it reduces database processing time as shown in Fig. 1. Further, there are many methods of encryption for the security purposes [5].

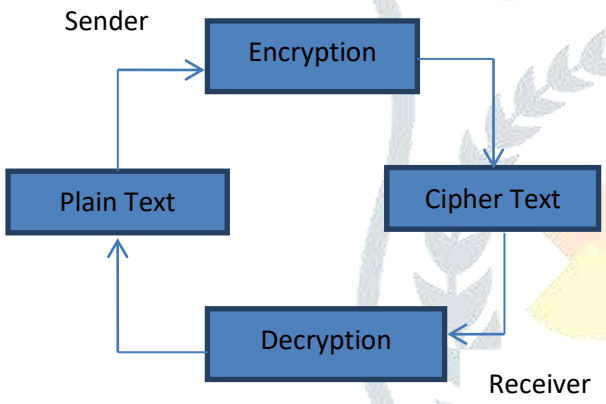


Fig. 1: Encryption and Decryption Process

### E. *Key Management*

Encryption and Decryption keys are needed for encryption purposes, but management of keys is the major issue in cloud environment. It is also not secure to store keys in the cloud. To store single key is a good idea, but it becomes complex task to store these keys. Moreover, separate database is needed to store all the keys. But shifting of database on cloud is also a tough job. If we add another hardware and software resources, then it will increase cost of the system. So, 2- level encryption is the only solution for the secure key management. This can be very helpful to store encryption keys in cloud.

### F. *Data Splitting*

Another alternative of the encryption is data splitting. It is very fast process as compared to the encryption. The main idea behind it is that split of the data over the number of hosts that are non-communicable.Each and every time a user requests its data back, he must have access to both of the service providers to recollect his original data. No doubt, it is very fast technique but it has its own security issues [7].

### G. *Multi-Tenancy*

Variety of resources and services in cloud environment are shared among various applications at different geographic locations. To solve the issues of resource limitations and cost is the main concern of the cloud. This is done to solve the issues of resource scarcity and to eliminate cost that is the main purpose of the cloud. But the distribution of the resources of an organization gives birth to confidentiality issues. To some extent, in order to keep confidentiality alive, these systems and applications must be inaccessible [8]. Else it will be difficult to check the data flow of the system. Data and applications in a cloud may be deposited on simulated and actual hardware as well, there is a need to resolve both of the issues. There are chances that one virtual machine is hosting a malicious application, if these are stored virtually can affect the performance of other machines. If these are stored on actual hardware, there may be security issues because of multi-core processing.

As we discussed above, there are various security issues in cloud computing. To overcome these issues solutions of problems are also there. In Fig. 2 we have discussed issues and alternative solutions of the problem.

## III. ALGORITHMS FOR SECURITY OF BIG DATA IN CLOUD

There are various algorithms to enhance security of big data in cloud as shown in Fig.2. These algorithms are applied according to the requirements and security level. In this section we will discuss these algorithms one by one.

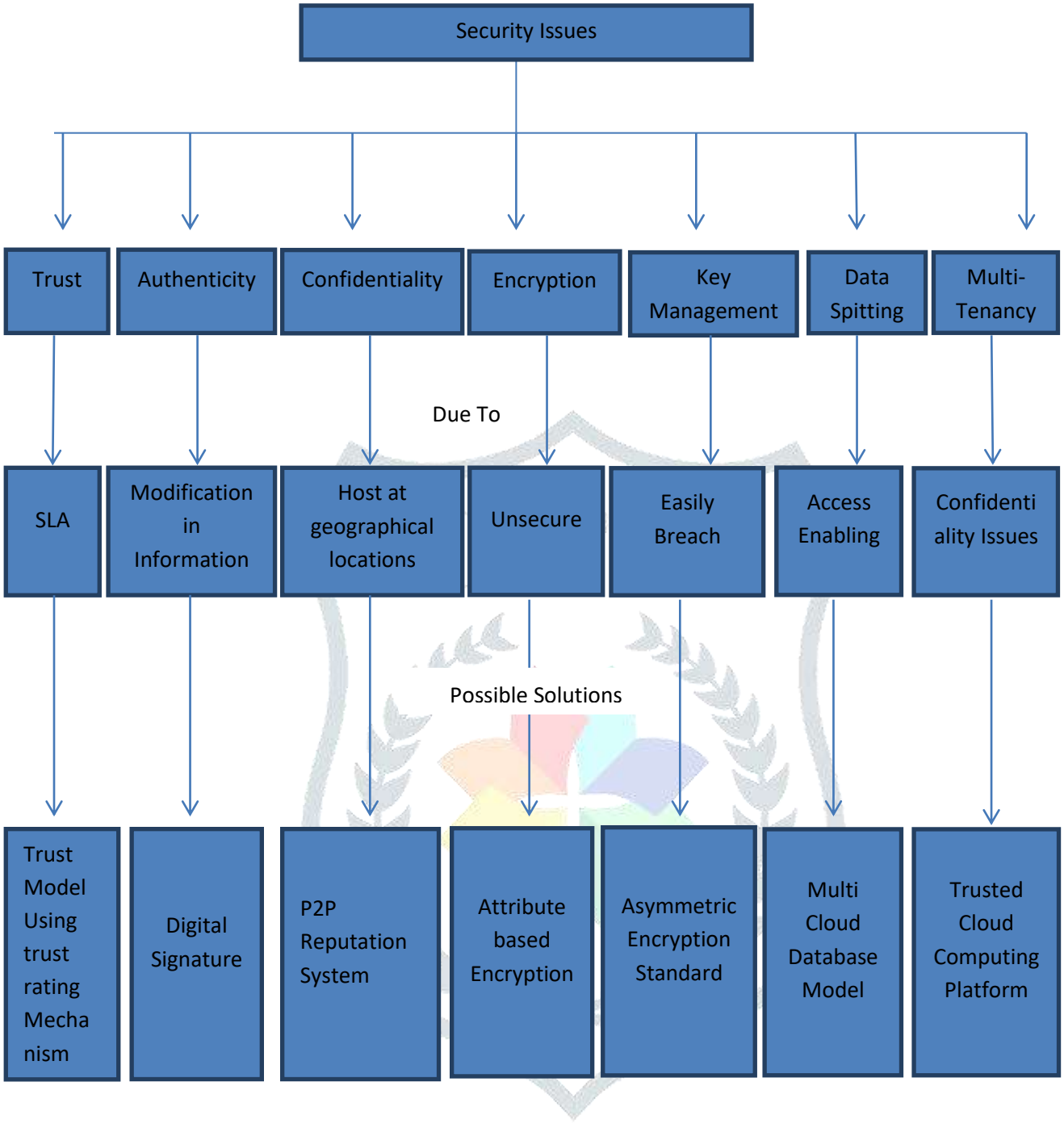


Fig.2: Classification of Security Issues and Its Possible Solutions

### A. *Trust Model Using Trust Rating Mechanism*

The trust rating algorithm basically depends upon the judgment of user's activities with their peers who registered in the same period of time. The main parameters which are considered for comparison purposes are availability, popularity, and user participation. Weights are given to the each user selection. The weighted sum of these factors is trust score. As a result, a user-adaptive access control can be attained [6]. The score determines if a user can read another user's blog. If the user's trust score meets a certain threshold set by the blog owner than a user can view a blog only.

### B. *Digital Signature*

For secure authentication process a mathematical scheme is used known as digital signature. The main purpose of this signature is that message is created by the sender only.A sender never deny about the sent message. To secure integrity, that message should not be altered. Digital signatures are usually used for financial transactions, software distribution and in some cases it is significant to identify forgery or tampering [9]. A Digital Signature is a combination of 0 & 1s created using crypto algorithms.

### C. *P2P Reputation System*

In this model, network is divided into four categories [12]. These categories are Peer, Honest Peer, Selfish, Evil Peer and Malicious Peer. In Honest Peer, Peers initiate's only good transactions. Selfish peers are free riders. He blocked all the transactions and refuses to give rating to its users. Malicious peers can generate any kind of transactions. Their behavior may damage all the system. Evil Peer try to gain good reputations by grouping people, he always try to give good rating to other evil.

### D. *Asymmetric Encryption Standard(AES)*

AES cryptographic algorithm works in rounds that depend upon the size of input data. As AES perform on 128 bit data to encrypt, so in this case 10 rounds are performed to encrypt the input data. In each single round, there are four modules performed to encrypt the data and the result of one round is given to next round for their working [14].

a) Sub Byte: - In the first module of each round first S-Box operations is performed. According to the S-box (substitution box) each byte of input state is replaced as shown in Fig. 3. Independent transformation work on single cell of the state.

| | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | A | B | C | D | E | F |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 63 | 7C | 77 | 7B | F2 | 6B | 6F | C5 | 30 | 01 | 67 | 2B | FE | D7 | AB | 76 |
| 1 | CA | 82 | C9 | 7D | FA | 59 | 47 | F0 | AD | D4 | A2 | AF | 9C | A4 | 72 | C0 |
| 2 | B7 | FD | 93 | 26 | 36 | 3F | F7 | CC | 34 | A5 | E5 | F1 | 71 | D8 | 31 | 15 |
| 3 | 04 | C7 | 23 | C3 | 18 | 96 | 05 | 9A | 07 | 12 | 80 | E2 | EB | 27 | B2 | 75 |
| 4 | 09 | 83 | 2C | 1A | 1B | 6E | 5A | A0 | 52 | 3B | D6 | B3 | 29 | E3 | 2F | 84 |
| 5 | 53 | D1 | 00 | ED | 20 | FC | B1 | 5B | 6A | CB | BE | 39 | 4A | 4C | 58 | CF |
| 6 | D0 | EF | AA | FB | 43 | 4D | 33 | 85 | 45 | F9 | 02 | 7F | 50 | 3C | 9F | A8 |
| 7 | 51 | A3 | 40 | 8F | 92 | 9D | 38 | F5 | BC | B6 | DA | 21 | 10 | FF | F3 | D2 |
| 8 | CD | 0C | 13 | EC | 5F | 97 | 44 | 17 | C4 | A7 | 7E | 3D | 64 | 5D | 19 | 73 |
| 9 | 60 | 81 | 4F | DC | 22 | 2A | 90 | 88 | 46 | EE | B8 | 14 | DE | 5E | 0B | DB |
| A | E0 | 32 | 3A | 0A | 49 | 06 | 24 | 5C | C2 | D3 | AC | 62 | 91 | 95 | E4 | 79 |
| B | E7 | C8 | 37 | 6D | 8D | D5 | 4E | A9 | 6C | 56 | F4 | EA | 65 | 7A | AE | 08 |
| C | BA | 78 | 25 | 2E | 1C | A6 | B4 | C6 | E8 | DD | 74 | 1F | 4B | BD | 8B | 8A |
| D | 70 | 3E | B5 | 66 | 48 | 03 | F6 | 0E | 61 | 35 | 57 | B9 | 86 | C1 | 1D | 9E |
| E | E1 | F8 | 98 | 11 | 69 | D9 | 8E | 94 | 9B | 1E | 87 | E9 | CE | 55 | 28 | DF |
| F | 8C | A1 | 89 | 0D | BF | E6 | 42 | 68 | 41 | 99 | 2D | 0F | B0 | 54 | BB | 16 |

Fig.3: Example of S-Box for AES

Resultant matrix is generated from the matched combination of each byte with the corresponding value in S-Box. This resultant matrix is thus further processed by shift row.

b) Shift Row: - On each row of the matrix, this transformation is applied. Length of the block is used to measure the offsets of each row as shown in Fig. 4. In case of Decryption, the rows are cyclically shifted left with offset equal to number of columns of State minus the offset for Encryption [14].

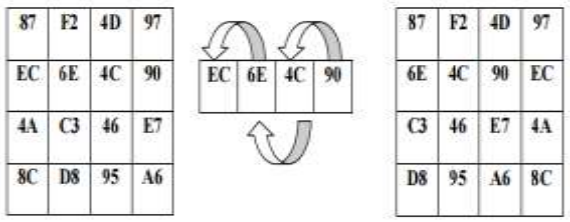


Fig.4: Row Transformation

c) Mix Columns: - After row shifting, Mix col operation is performed. Each column of matrix is taken and then multiplied with a matrix as depicted in Fig. 5.

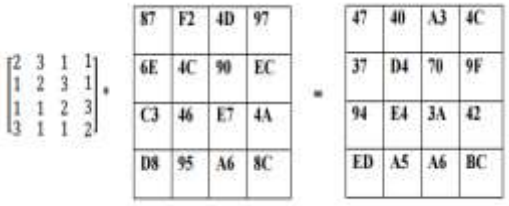


Fig. 5: Mixed Columns

d) Add Round Key ():- In this transformation, 128 bits key is XORed with 128 bits of the round key. As shown in Fig. 6,this operation is work as column wise with each 4 bytes of the state.

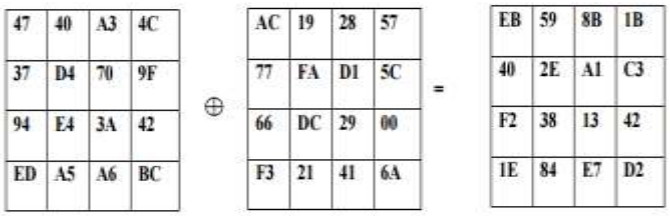


Fig.6: Add Round Keys

These all above steps are performed in each round of AES algorithm and the result of one round is given as an input to another round. After all the 10 Rounds (For 128 bit data), the result of encryption process is found that is a cipher text and reverse of this process is called decryption and user again got plain text.

### E. *Attribute Based Encryption (ABE)*

ABE is an algorithm based on the attributes and uses number of keys for encryption and decryption. Cipher text

and secret key of user depends on the attributes. Moreover, the decryption of a cipher text is promising, if user key matches with the cipher text attributes [13]. Decryption takes place only when the number of matching is at least a threshold value d. Collusion-resistance is vital security feature of Attribute-Based Encryption. The main problem in this algorithm is that data owner has to use every key for encryption. Further it is divided into two categories:

1. Key Policy ABE
2. Cipher Text Policy ABE

In KP-ABE, a set of attributes is related with cipher text and the user's decryption key is linked with a monotonic access tree structure. On the other hand, in CP-ABE,every user is related with a set of attributes. His secret key is generated based on his attributes. In this case, threshold access structure is specified by the encryptor according to the interested attributes.This message is then encrypted established on access structure such that only those whose attributes satisfy the access structure can decrypt it. With CP ABE technique, data is secure and confidential against attacks due to encryption [13].

TABLE I
COMPARISON OF ACCESS MODEL

| **Parameters** | **Key Policy ABE** | **Cipher Text ABE** |
|---|---|---|
| Efficiency | Excellent Performance for Broadcast system | Average, Not Efficient for modern environment |
| Fine grained Access Control | High if and only if 2-level encryption | Average, Complex Access structure |
| Collision | Good | Good |
| Computational Overhead | Most | Average |

### *F. Multi Cloud Database Model*

This model not only maintains encryption, but it is maintained with the help multi-cloud service providers and secret key sharing mechanism. Various types of data queries are permitted by these types of algorithms to prevent risk of attack[15].Therefore, data is replicated among a predefined number of cloud service providers (CSP) using Shamir's algorithm. Data and its operations are handled and controlled with the help of DBMS between cloud and cloud service provider.Moreover, it produces the polynomial values, tackles the user's queries to each cloud, recreates the result and finally directs it back to the client.

### *G. Trusted Cloud Computing Platform*

Trusted cloud computing platform (TCCP) offers an environment by spreading the concept of trusted platform to an entire IaaS backend for a closed box execution. The TCCP promises the confidentiality and the reliability of a user's VM and favors a user to control whether or not the IaaS enforces security. TCCP does the job of all the trusted nodes as one entity. Therefore the network has number of entities and create cluster among those entities.If one entity turn out to be the failure than whole approach can't perform smoothly [15].

TABLE II
COMPARISON WITH OTHER METHODS

| **Algorithms/ Parameters** | Security | Robustness | Time Complexity | Space Complexity |
|---|---|---|---|---|
| Trust Model using Trust rating Mechanism | Provide best security in case of SLA | Average | More Complex | Less Complex |
| Digital Signature | Provide best security in case of authentication | Good | Less Complex | Less Complex |
| P2P Reputation System | Provide best security in case of confidentiality | Average | Large Complex | Large Complex |
| Attribute Based Encryption | Provide best security in case of encryption | Strong | Less Complex | Less Complex |
| Advance Encryption Standard | Provide best security in case of Breeching | Good | Large Complex | Large Complex |
| Multi-cloud Database Model | Provide best security in case of access enabling | Average | More Complex | More Complex |
| Trusted Cloud Computing Platform | Provide best security in case of confidentiality | Average | More Complex | Large Complex |

## IV. CONCLUSION AND FUTURE SCOPE

In this paper, we have analyzed various security issues for big data in cloud. Meanwhile, it is also concluded that these issues are responsible to degrade the performance of the network. Therefore, to overcome these issues, their possible solutions have been mentioned in this paper. We mainly focused on the Asymmetric Encryption Standard and Attribute Based Attributes to enhance security of big data in cloud. Furthermore, according to our survey, KP-ABE is the best technique to enhance cloud security in terms of efficiency and computational overhead. In future, various attacks on big data in cloud can be discussed and they can be prevented using ABE technique.

## REFERENCES


[1] Li, Wenjuan, and Lingdi Ping. "Trust model to enhance security and interoperability of cloud environment." In Cloud Computing, pp. 69-79. Springer Berlin Heidelberg, 2009.

[2] Ko, Ryan KL, et al. "TrustCloud: A framework for accountability and trust in cloud computing." Services (SERVICES), 2011 IEEE World Congress on. IEEE, 2011.

[3] Weiyi Shang, Zhen Ming Jiang, HadiHemmati, Bram Adams, Ahmed E. Hassan, Patrick Martin "Assisting Developers of Big Data Analytics Applications When Deploying on Hadoop Clouds",IEEE 978-1- 4673-3076-3/13 IEEE,2013.

[4] Nada Elgendy, Ahmed Elragal, "Big Data Analytics: A Literature Review Paper", ICDM, LNAI 8557, pp. 214–227, 2014.

[5] Li, Wenjuan, and Lingdi Ping. "Trust model to enhance security and interoperability of cloud environment". In Cloud Computing, pp. 69-79. Springer Berlin Heidelberg, 2009.

[6] Wooten, Ryan, et al. "Design and implementation of a secure healthcare social cloud system." Cluster, Cloud and Grid Computing (CCGrid), 2012 12th IEEE/ACM International Symposium on.IEEE, 2012.

[7] M. Alhamad, "Conceptual SLA Framework for Cloud Computing", Accepted for IEEE DEST 2010 on 15 March 2010.

[8] Alhamad, Mohammed, Tharam Dillon, and Elizabeth Chang. "Sla-based trust model for cloud computing." Network-Based Information Systems (NBiS), 2010 13th International Conference on.IEEE, 2010.

[9] Roschke, Sebastian, Feng Cheng and Christoph Meinel. "Intrusion detection in the cloud". Dependable, Autonomic and Secure Computing, 2009. DASC'09. Eighth IEEE International Conference on IEEE, 2009.

[10] Santos, Nuno, Krishna P.Gummadi, and Rodrigo Rodrigues "Towards trusted cloud computing". Proceedings of the 2009 conference on Hot topics in cloud computing, 2009.

[11] Onwubiko, Cyril. "Security issues to cloud computing." Cloud Computing, Springer London, 2010. 271-288.

[12] Sandeep Kumar, ChanderDiwaker, Amit Chaudhary, "Reputation System in Peer-to-Peer Network: Design and Classification" Journal of Global Research in Computer Science, Volume 2, No. 8, September 2011.

[13] Minu George, Dr. C.SureshGnanadhas, Saranya.K, "A Survey on Attribute Based Encryption Scheme in Cloud Computing", International Journal of Advanced Research in Computer and Communication Engineering Vol. 2, Issue 11, November 2013.

[14] TannuBala and Yogesh Singla, "Comparative Analysis of Parallel AES Algorithm with Pipelined AES Algorithm", International Journal of Computer Science Trends and Technology (IJCST) – Volume 3 Issue 6, Nov-Dec 2015 ISSN: 2347-8578.

[15] DivyeshYogan and Pooja Kose, "Trusted Cloud Computing Platform into Infrastructure as a Service Layer to Improve Confidentiality and Integrity of VMs", International Journal of Computer Applications (0975 – 8887) Volume 131 – No.7, December2015.

[16] Zhidong Shen, Li Li, Fei Yan, and Xiaoping Wu, "Cloud Computing System Based on Trusted Computing Platform," Intelligent Computation Technology and Automation (ICICTA), IEEE International Conference on Volume: 1, pp. 942-945. China. 2010.

[17] Xiao-Yong Li, Li-Tao Zhou, Yong Shi and Yu Guo, "A Trusted Computing Environment Model in Cloud Architecture," Proceedings of the Ninth International Conference on Machine Learning and Cybernetics, 978-1- 4244-6526-2. Qingdao, pp. 11-14. China. July 2010.

[18] Qingsong Wei, BharadwajVeeravalli, Bozhao Gong, Lingfang Zeng and Dan Feng, "CDRM: A Cost-Effective Dynamic Replication Management Scheme for Cloud Storage Cluster," 2009 IEEE International Conference on Cluster Computing (CLUSTER), pp. 188-196, 2010.

[19] Kai Hwang, Sameer Kulkareni, and Yue Hu, "Cloud Security with Virtualized Defense and Reputation-Based Trust Mangement," 2009 Eighth IEEE International Conference on Dependable, Autonomic and Secure Computing (DASC '09), pp. 717-722, 2009.

[20] Zhao-xiong Zhou, He Xu, and Suo-ping Wang, "A Novel Weighted Trust Model based on Cloud," AISS: Advances in Information Science and Service Sciences, Vol. 3, No. 3, pp. 115- 124, April 2011.

[21] Yunchuan Sun, Junsheng Zhang,2Yongping Xiong and Guangyu Zhu, "Data Security and Privacy in Cloud Computing" Hindawi Publishing Corporation International Journal of Distributed Sensor Networks Volume 2014, Article ID 190903, 9 pages