Review on Data Classification Mechanisms Used In Cloud Computing.

S.Ramalakshmi[1], Dr.V.Vallinayagi[2]

[1]Assistant Professor& Research scholar, Reg No :18221262162004,
Department of Computer Science,
Sri Sarada College for Women, Tirunelveli–11.
[2]Head & Associate Professor, Department of Computer Science,
Sri Sarada College for Women, Tirunelveli–11.
Manonmaniam Sundaranar University, Abishekapatti, Tirunelveli – 627 012.
Email: lakshmigana2011@gmail.com

ABSTRACT

Cloud computing is a type of internet based computing which means storing and accessing data through the internet platform instead of personal computer. Cloud provides number of software, infrastructure,platform and storage services to users. Sheltering user data is the challenging issue in the cloud environment as data is stored and maintained by third party service providers.To address this issue numerous security procedures are followed.Encryption is a peculiar technique used in cloud to protect user data.Encrypting entire data without examining its security needs will lead to excess amount of storage and additional processing power. Outcome of this problem is a classified encryption method [1].According to this method data is classified based on its security level and application of encryption algorithms. Sensitive data needs higher level of encryption techniques whereas basic data does not require these encryption techniques.Various classification algorithms are used such as“Decision tree”,” Random forest”, “KNN” (K Nearest Neighbor), “Naïve Byes” and “C4.5” are some of them.These algorithms follow the supervised learning methods which means these algorithms analyze the new training data with previously labeled data and gives proper output. Main objective of this data classification in cloud is to aim for better security and coherent usage of memory [2]. In this paper, detailed study of these data classification methods used in cloud environment are exposited.

KEY WORDS: Encryption algorithms, Classification algorithms, Decision tree, Random forest, KNN.

1. INTRODUCTION

Cloud computing refers to accessing services and software applications through online. The word cloud means metaphor for internet. Cloud offers services that are available to users on demand basis through the internet. Cloud provides scalable and reliable access to resources, software applications, platform services and storage. Storage is an efficient service provided by cloud to its users. All these services are maintained and controlled by cloud service providers. Important services provided by cloud are listed below.

1. SAAS →Storage As A Service
2. SAAS →Software As A Service

3. PAAS→Platform as a Service
4. IAAS→Infrastructure As A Service

Storage as a Service means which is a service cloud offers that user can store, read and retrieve data via an internet by using utility computing.

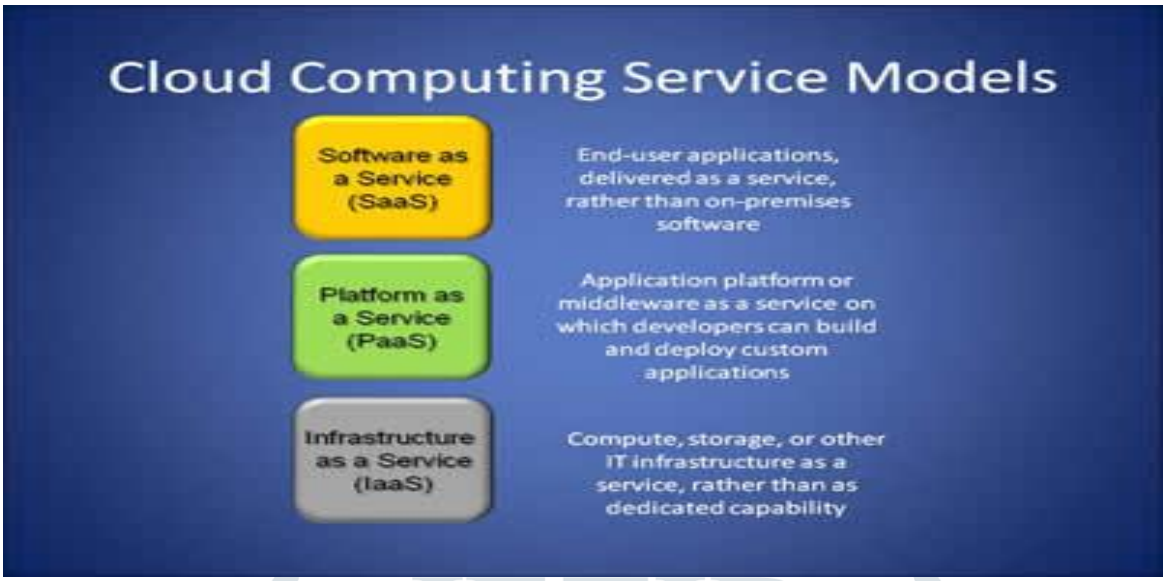

FIGURE 1 – CLOUD SERVICE MODELS (ipfiles.wordpress.com)

User data are stored and maintained by a third party cloud service providers, so achieving security is a biggest challenge here. When a user migrating their data to cloud have to be consider the following risks , Intruder and cyber attacks, Legal liability, Data theft, Data loss Encrypting data is the first head solution to protect user data from these risks but encrypting the entire data without examining its security needs will lead to excess amount of storage, unnecessary maintenance issues. Classified encryption is a best solution to address this issue This paper deeply discuss about some of the classified encryption mechanism used in cloud.

2. RELATED WORK

A secure cloud computing model based on manual data classification, classified data are encrypted by using different encryption algorithms. It minimizes the overhead and the processing time essential to secure data through using different security mechanisms [1]. Major growth and major drawbacks of cloud computing was discussed. Cloud has good solution to increase productivity in much area such as cost effectiveness, reduced time and efforts. The major issues are computability and interoperability [2].To secure the data in the cloud database server cryptography is the essential method. Analyzed various symmetric and asymmetric algorithms used in cloud [3].Different cryptographic security protocols used for cloud computing are discussed. Solutions for few of the security issues in cloud are introduced that are, i) Intrusion detection system, ii) Cloud computing security gateway ,these two techniques that are discussed here[5]. Avoiding third party auditors is the main concept here, it provides the customer, securing their data by their self in a very simple way by using an existing hash generating algorithm [6].Cloud data is transforming to the unknown destination, so there is a need for effective security mechanism. So Hybrid Cryptographic Algorithms are proposed [7].Data classification means here to separate sensitive and non sensitive data by using KNN classification algorithm. RSA encryption algorithm is used to encrypt the sensitive data and the non sensitive public data is directly stored in the virtual machine [8].Classification of data is achieved through machine learning techniques by using K Nearest Neighbor algorithm. Data is classified as public and private data. Private data is encrypted and

then both of public and private data is stored in cloud storage [9]. Comparative analysis on the performance of selected security algorithms in cloud computing They proved blowfish exhibited better performance than RSA and AES algorithms [10]..Authentication, Authorization and encryption are the three main goals to achieve in cloud. 100GB which are going to deploy over the Cloud and out of which 15% of data is sensitive which require more security. So encrypting the whole data with same level of security and with same key size encryption then it is not feasible in terms of processing time because it takes more time and except for data other than 15% is waste of time [11]. A model that classifies the data according to its security parameters. The performance of the existing KNN is improved by appending it with ensemble learning technique [12].

3. ENCRYPTION WITH DATA CLASSIFICATION

ENCRYPTION: Encryption means encoding a message or information by using various encryption algorithms that unauthorized user cannot access the message.. Encrypted data is called cipher text. Symmetric and asymmetric key encryptions are the main two categories. In recent trends data is stored and maintained in cloud storage, privacy is the main concern to think of. The most popular solution for this is encrypting data .Encrypting the entire data without considering its security needs will leads to extra processing time, vast amount of storage needs, so classification of data based on its sensitivity and encryption based on the classification will help this situation for betterment. Amount of storage and processing time will be decreased when classified encryption concept is applied.

DATA CLASSIFICATION: Classifying user data based on its security needs is the process to be done before encryption. To achieve this classification some algorithms are used. These classification algorithms are fall under two main categories that are Supervised and Unsupervised learning algorithms.

SUPERVISED OR MACHINE LEARNING: Here data is labeled and algorithm learns to produce the output data from the input data. K Nearest Neighbor (KNN), Naive Bayes, Support Vector Machine, Random Forest, Decision Tree are some of the machine learning algorithms.

UNSUPERVISED LEARNING: Here data is unlabeled and the algorithm learns to immanent the basic structure from the input data. Cluster analysis, K means clustering, Artificial neural network are some examples for unsupervised learning algorithms .Below some of the classified encryption of data used in cloud are discussed.

3.1 Security Perturbation in cloud using level classification

Machine learning technique for classification of data, this mechanism is rarely used in cloud computing. Here supervised learning technique is used and data are labeled with class name. K Nearest Neighbor is the classification algorithm used to classify the data. Additionally improved bagging technique also used with KNN algorithm. Bootstrap aggregation is used to aggregate the classifiers to improve the performance. It chooses a training set of size N for classifier K+1 by selecting examples from the original N training examples. To improve accuracy and reduce error rate they combined these techniques. Here data is classified into four types,

i) Top Secret
ii) Secret

iii) Confidential
iv) Unclassified.

Bell – lapadula model rule followed to check the confidentiality of the classified data. After this classification by using KNN algorithm the encryption is performed based on the classification. Top level secret data RSA asymmetric cryptographic algorithm is used to achieve highest level of security. Blow fish algorithm is used for confidential data. They classify the data into various classes on the level of sensitivity and to secure sensitive data hybrid encryption algorithms are used[10].

3.2 KNN Classifier for Data Confidentiality in Cloud

Classification of data by using machine learning technique is used here. Supervised learning algorithms applied for data classification. Data is classified into two types that are

i) Non Sensitive
ii) Sensitive

Sensitive Data: Personal data, financial records, Business material, Legal data, Government data are labeled as sensitive data and others are non sensitive.

Non Sensitive Data: Non sensitive data is considered as public data.

K Nearest Neighbor Classification algorithm is used to classify the data based on the labels. RSA encryption algorithm is used to encrypt the Sensitive data and at the end Sensitive data and Non sensitive data are stored under different virtual machines. This research proven for efficient usage of storage under cloud server. The CloudSim simulator was used for simulation purposes. The Virtual Machine Manager (VMM) is used for manage and allocate virtual machines to cloudlets. IAAS and PAAS properties for cloud simulation was discussed and mentioned by tables. Data selected for this process is employee's records. This research proposed a confidential based data classification model for cloud computing [11].

4. RESULT ANALYSIS

Analysis of level classification of data, classification and encryption algorithms used to fulfill the security needs of data are listed in the below table.

Table 1 – Classification Analysis

Authors	*Data Set*	*Classification Algorithm*	*Level of Classification*	*Encryption Algorithms*
Er.Vikram Dhiman Er.Himakshi Er.Ankushdeep Kaur Dr Prof Manoj Kumar	Various datasets for different classes	KNN Classifier + Bootstrap Aggregation	a. Top Secret b. Secret c. Confidential d. Unclassified	a. RSA (Top Secret) b. Blow fish (Confidential)
Munwar Ali Zardari Low Tang Jung Nordin Zakaria	Employee dataset	KNN Classifier	a. Non Sensitive b. Sensitive	RSA (Sensitive)
Radha Patel Satish Dehariya	Personal data (images,videos,text documents)	Manual Claassification	a. Basic b. Sensitive c. Highly Confidential	a. AES-128 b. AES-256 c. AES-512

From the review of different papers based on level classification of data KNN is the most commonly used machine learning classification algorithm ,it classifies labeled data and produce output data from that labeled data.RSA encryption algorithm is used to fulfill the security needs of the highly confidential data.

5. CONCLUSION AND FUTURE SCOPE

Data classification method enhances the security of data and gives different levels of authentication for various classified data. Manual classification and algorithm based classification techniques are used for data classification. Hybrid encryption algorithms for different level of data are applied. Finally security level is improved; amount of storage and processing time is decreased.

REFERENCES

[1] Amol C. Adamuthe, Vikram D. Salunkhe, Seema H. Patil ,Gopakumaran T. Thampi , "Cloud Computing– A market Perspective and Research Directions", I.J. Information Technology and Computer Science, 10, 42-53, , 2015.

[2] Prof. Dr. Christof Weinhardt,Arun Anandasivam,Dr. Benjamin Blau,Nikolay Borissov,Thomas Meinl,Dr. Jochen Stößer,"Cloud Computing – A Classification, Business Models, and Research Directions", DOI 10.1007/s12599-009-0071-2

[3] Hatem M. Abdul Kader, Mohie M. Hadhoud, Salah M El-Sayed, Dia SalamaAbdElminaam, "Performance Evaluation Of New Hybrid Encryption Algorithms To Be Used For Mobile Cloud Computing",International journal of technology enhancements and emerging engineering research, vol 2, issue .4 63 ISSN 2347-4289

[4] R. Bala Chandar, M. S. Kavitha and K. Seenivasan,"A proficient model for high end security in cloudcomputing", Ictact journal on soft computing, volume: 04, issue: 02, January 2014.

[5] Raj Kumar, "Research onCloud Computing Security Threats using Data Transmission", International Journal of Advanced Research in Computer Science and Software Engineering. Volume 5, Issue 1, January 2015.

[6] Smita Parte, Noumita Dehariya, "Cloud Computing: Issues Regarding Security, Applications andMobileCloud Computing", International Journal of Advanced Research in Computer Science and Software Engineering. Volume 5, Issue 3, March 2015.

[7] Changyou Guo, and Xuefeng Zheng,"The Research of Data Security Mechanism Based on Cloud Computing ", International Journal of Security and Its Applications Vol. 9, No. 3 (2015), pp. 363-370.

[8] Rachna Arora, Anshu Parashar,"Secure User Data in Cloud Computing Using Encryption Algorithms ", International Journal of Engineering Research and Applications (IJERA). Vol. 3, Issue 4, pp.1922-1926, Jul-Aug 2013.

[9] Randeep Kaur, Supriya Kinger,"Analysis of Security Algorithms in Cloud Computing ", International Journal of Application or Innovation in Engineering & Management (IJAIEM). Volume 3, Issue 3, March 2014.

[10]Er.Vikram Dhinman ,Er.Himakshi,Er.Ankushdeep Kaur,Dr.Prof Manoj Kumar,"Pragmatic approach to conquer security perturbation in cloud computing using level classification",2017 2nd International Conference for Convergence in Technology(12CT).

[11]Faraz Fatemi Moghaddam,Aida Majd,Mohammad Ahmadi,Touraj Khodadadi,Kasra Madadipouya,"A Dynamic classification Index to Enhance Data Protection Procedures in Cloud –based Environments",2015 IEEE 6th Control and System Graduate Research Colloquium,Aug 10-11,UiTM,Shah Alam,Malaysia.

[12] Radha Patel, Sathish Dehariya, "Secure Model for Cloud Computing by Using Data Classification Methodology", International Journal of Innovative Research in Computer and Communication Engineering, Vol 4, Issue 12, December 2016.