ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 1, January 2017

Enhancing Cloud Security with Automatic Data Classification and Appropriate Encryption Algorithms

Ramalakshmi.S[1], Rexy.J[2]

Research scholar, Dept. of Computer Science, ST.Xaviers College, Tirunelveli, Tamilnadu, India[1]

Assistant Professor, Dept. of Computer Science, ST.Xaviers College, Tirunelveli, Tamilnadu, India[2]

ABSTRACT: Cloud computing is an internet based computing, the word "cloud" is used as a metaphor for the internet. Cloud computing changes the trends to store and retrieve data.Users can store their data in the remote servers and can retrieve it anywhere, anytime provided an internet data connection is readily available. Now a days the data is considered as an intellectual asset for the users. So security is an important aspect in this regard. In a large dataset there may be a smaller portion of data which needs security. The existing solution for this problem is manually classifying the data and encrypting the data based on the classification. Manual classification of a larger database is a time consuming process. So in this proposed work, classification is done automatically by applying proper algorithms and data is encrypted based on the classification. This method reduces the processing power, memory usage and the time consumed.

KEYWORDS: Cloud computing, Data classification, Encryption, Classification algorithm.

I.INTRODUCTION

Cloud computing is one of the emerging technology in the present scenario. Storing and retrieving of data are done under the remote server. Data may contain financial information, business information and personal data. Thesedata are stored and managed by third party service providers. Amazon (AWS), IBM Clouds, Google are some of the cloud service providers.

The only concern being is the security issues which causes most users to think on using cloud technology. According to KMPK cloud security report 2014, 53% of respondents identified data loss and data privacy as their top cloud challenges. The most popular solution for this is encrypting data.

Sometimes a larger database may contain only small amount of sensitive datawhich needs to be highly protected and otherdata are considered as basic datawhich needs only basic encryption methods. The existing solution for this problem is manually classifying the data and encrypting the data based on the classification.

In a large database manual classification takes more time to classify all data. In this proposed work, classification is done automatically by using proper algorithms and according to the classification different encryption algorithms are used to encrypt the data. This method provides a faster classification and a secure encryption for data to be stored in cloud or any other online storage means.

II.RELATED WORK AND EXISTING SOLUTIONS

In [1] Major growth and major issues of cloud computing was discussed. Cloud has good solution to increase productivity in many concerns such as cost effectiveness, reduced time and efforts. The major issues are security, privacy, computability and interoperability.

In [2] A classification, business models and research directions of cloud computing are discussed, says that there is not more transparency available at which location the data is stored. So security and trust is the major research issue.

In [3] Cloud data is moving to the unknown destination, so there is a need for effective security mechanism. SoHybrid Cryptographic Algorithms are proposed.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 1, January 2017

In [4] Data integrity, data loss and secure data access is the major issues in cloud, so here the data has been converted into more scalable and flexible form to add more security at access levels.
In [5] Three ways are introduced to achieve data security, that are i) Data transmission, ii) Data isolation, iii) Data wiping.IP sec,VPN and SSL are able to incorporated within cloud to improve the data security.
In [6] Solutions for some security issues in cloud are introduced that are i) Intrusion detection system, ii) Cloud computing security gateway.
In[7] Proposed additional securities for a secured cloud computing such as physical security, cloudOS security, data security, virtual cluster security, data security, SaaS/paas security.
In [8] Different encryption algorithms used in cloud computing are compared with each other.
In [9] Different security algorithms used in cloud was analysed.
In [10] Protect the data in the cloud database server cryptography is one of the method. Analysed various symmetric and asymmetric algorithms used in cloud.
In[11] Data classification used to achieve high data security and effective use of encryption algorithms .Data stored in the cloud are manually classified and based on this classification various encryption algorithms are used.
In [12] Survey about classification techniques used in cloud computing and for security AES, hybrid encryption algorithm, homomorphic encryption schemes are used.
In [13] Data classification technique is used to improve cloud security. Data are classified under three different criteria Access control, Content, Storage. According to this classification security considerations can be applied.
In [14] proposed 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 needed to secure data through using different security mechanisms.

III.PROPOSED WORK WITH AUTOMATIC DATA CLASSIFICATION

III.1. METHODOLOGY:

Main objective of this work is to achieve higher security with the use of automatic data classification.

Model Database	Test Name	Haematology Fields	Bio-chemistry Fields
Patient Details(Blood Test Results)	Haematology Bio-chemistry	1.TRBC Count 2.TWBC Count 3.Haemoglobin 4.Platelets Count	1.Blood Glucose(F) 2.Blood Glucose(PP) 3.Creatinine

DATABASE:

The patients detail database contains the blood test result values of the patients. There are two main tests taken are named as Haematology and Bio-chemistry. Each test having various fields that are mentioned above. Thesedata are being classified into two levels that are

i) Sensitive level
ii) Basic level

The classification is done automatically by using Naive Bayes Classification Algorithm. Newly entered data are also automatically classified by using proper implementation of algorithm. After that these data are encrypted by using different encryption algorithms based on the classification.

SENSITIVE LEVEL:

In the above database each field has minimum and maximum range values. If the tested values of a patient fall beyond the normal range of each field, then the patient's record is considered as a sensitive record as it may contain confidential information about the disease. For example the field TRBC count having the minimum value 4 and the maximum is 6 if the patients predicted value is not under this range then that record is considered as a sensitive one. To encrypt this type of records RSA encryption algorithm is used.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 1, January 2017

BASIC LEVEL:
If all tested values of a patient are in the satisfied (normal) range then this type of record is considered as basic level records. In this case there is no disease related information is identified. So these kinds of records are basic one. To encrypt this type of records Unicode Encoding is used.

III.2. IMPLEMENTATION:

Procedure	Algorithm
Classification	Naïve Bayes
Sensitive level encryption	RSA-1024 bit algorithm
Basic level encryption	Unicode Encoding

Naïve Bayes Classification Algorithm:-
Gaussian Naive Bayes algorithm is used here for achieving automatic data classification. Minimum and maximum range values are applied for each field by using this algorithm. Based on this, the data is segmented by its rangesand then mean and variance are calculated.According to that,data are classified into two categories assensitive level and basic level.

RSA Encryption Algorithm :-(Sensitive Level)
RSA 1024 bit algorithm is used for sensitive level encryption.RSA algorithm was described by RonRivest,Adi Shamir and Leonard Adelman. It is an asymmetric block cipher encryption algorithm. Asymmetric means two different keys are used. One is public key that is known to all and the other is private key which is a secret one. Data are encrypted by using public key. It provides better security for sensitive data particularly used in online transactions.

Unicode Encoding :-(Basic Level)
Unicode encoding is used for basic level encryption. It can be implemented by different character encodings. The acronym UTF stands for Unicode Transformation Format. It uses 8 bit blocks to represent characters. Unicode encoding is ease of providing shared access to data. It has the advantages of interoperability and web compatibility.

IV.DESIGN GOALS

In this proposed work a model database named “Patient Detail Database” was taken for processing. This database contains 100 records.
Step 1: Import the database.
Patient detail database was imported. It contains the records of 100 patients

Step 2: Classify the records by using the predefined ranges. Gaussian Naïve Bayes algorithm is used here.
I) Classification of 100 records by using Gaussian Naïve Bayes classification algorithm.In the above database each field has minimum and maximum range values. If the tested values of a patient fall beyond the normal range of each field, then the patient’s record is considered as a sensitive record as it may contain confidential information about the disease. Here, from the 100 records 20 records are considered as a sensitive level records
II) If all tested values of a patient are in the satisfied (normal) range then this type of record is considered as basic level records. In this case there is no disease related information is identified. So these kinds of records are basic one. Here, from the 100 records remaining 80 records are considered as a basic level records.

Step 3:Sensitive records are encrypted by RSA encryption algorithm
Sensitive level 20 records are encrypted by RSA encryption algorithm.It is an asymmetric block cipher encryption algorithm. Asymmetric means two different keys are used. One is public key that is known to all and the other is

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

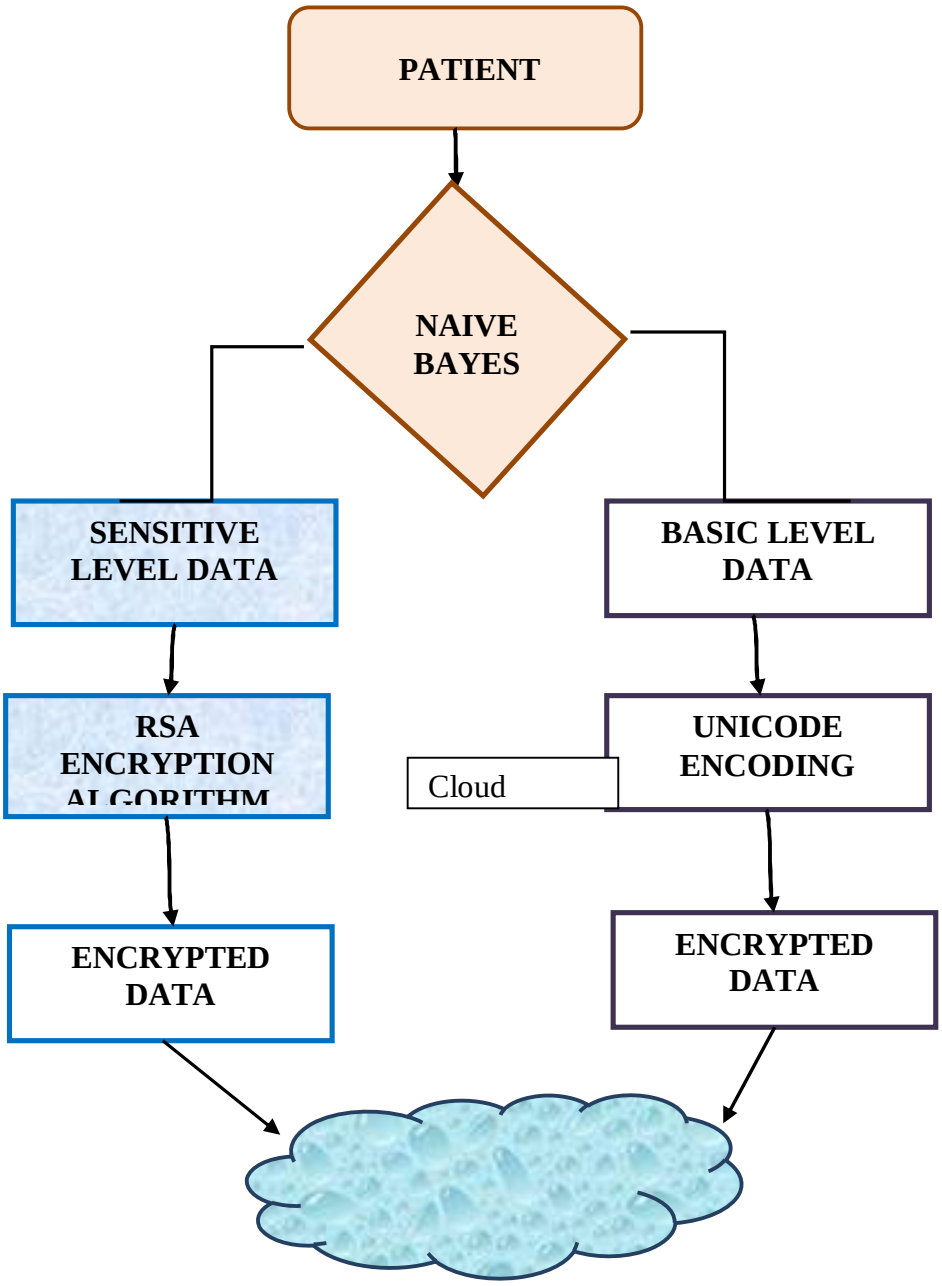
Vol. 5, Issue 1, January 2017

private key which is a secret one. Data are encrypted by using public key. It provides better security for sensitive data particularly used in online transactions. Memory occupied for this encryption is 720 bytes.

Step 4: Basic records are encrypted by Unicode Encoding.

Basic level 80 records are encrypted byUnicode encoding. It can be implemented by different character encodings. The acronym UTF stands for Unicode Transformation Format.

Following chart describes about the process step by step.



ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

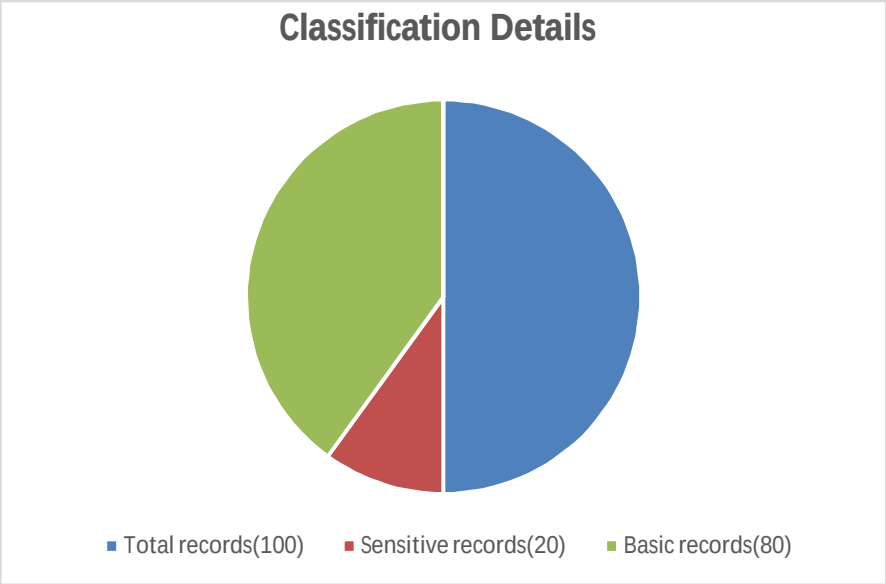
Website: www.ijircce.com

Vol. 5, Issue 1, January 2017

V. SIMULATION RESULTS

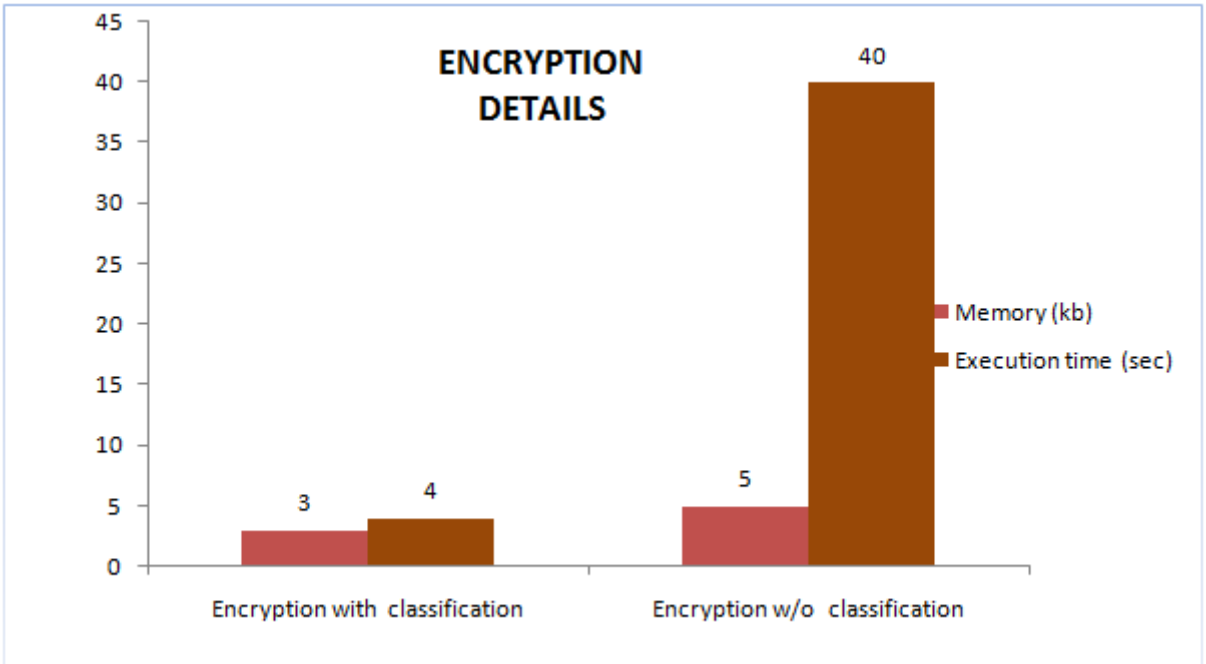
V.1.CLASSIFICATION RESULTS:

Following chart shows the classification details of sensitive and basic level data. Totally 100 records are taken for processing, from that 20 records are considered as sensitive and remaining 80 records are the basic level.



V.2.ENCRYPTION RESULTS

Following chart shows the performance comparison between encryption with classification vs encryption without classification. Encryption with classification consumes less memory and execution time comparing with the later one. It shows the importance of classification.



ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 1, January 2017

VI.CONCLUSION AND FUTURE SCOPE

In this proposed work, automatic data classification techniques are applied and based on the classification different level of encryption algorithms was used to provide security for data. The main scope of this work is to reduce processing power, efficient use of memory and also a cost efficient work. Naive Bayes classifier algorithm was used for classification, RSA asymmetric algorithm was used for encryption of sensitive level data,and Unicode Encoding is used for basic level encryption. As a part of our future work retrieving data from cloud and decryption will be performed and also classification of personal data including images, audios, and videos can be done automatically by applying proper classification methods.

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] Vishal R. Pancholi, Dr. Bhadresh P. Patel, "Enhancement of Cloud Computing Security with Secure Data Storage using AES", IJIRST – International Journal for Innovative Research in Science & Technology| Volume 2 | Issue 09 | February 2016 **.**
[11] Pankaj Pali,Saurabh Sharma," Security Model for Cloud Computing by using Data Classification Methodology", International Journal Of Innovative Research & Development. Vol 5 Issue 2, January, 2016.
[12]Prakash Sawle, Trupti Baraskar," Survey on Data Classification and Data Encryption Techniques Used in Cloud Computing",International Journal of Computer Applications (0975 – 8887) Volume 135 – No.12, February 2016.
[13]Rizwana Shaikha, Dr. M. Sasikumar,"Data Classification for achieving Security in cloud computing",Procedia Computer Science 45 (2015) 493 – 498.
[14] Lo'ai Tawalbeh, Nour S. Darwazeh, Raad S. Al-Qassas and Fahd AlDosari," A Secure Cloud Computing Model based on Data Classification", First International Workshop on Mobile Cloud Computing Systems, Management, and Security. Procedia Computer Science 52 (2015) 1153 – 1158.