ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 7, Issue 12, December 2018

Security to Data Access in Cloud Computing using Learning Techniques and Cryptanalytic Strategies

G.Sreelatha[1], Dr.A.Kanaka Durga.[2], T.Sandeep [3]

Assistant Professor, Department of IT, SCETW, Hyderabad, India [1]

Professor, Department of IT, SCETW, Hyderabad, India [2]

Assistant professor, Department of IT, SCETW, Hyderabad, India [3]

ABSTRACT: **Cloud computing** is a way to increase the capacity or add capabilities dynamically through remote servers hosted on the network to store, manage, and process data, rather than a local server or a personal computer. As technology is increasing, there's a rapid increase within the sensitive and crucial data. This leads to the necessity of securing the data of the users who are using cloud. In this paper, we proposed the data classification using machine learning technique KNN with modified algorithm (MKNN) to gain efficient results. The performance of MKNN learning technique improves prediction capability and classification accuracy of basic data, sensitive and highly confidential data. After classifying, to secure sensitive data, the cryptanalytic strategies are used. This helps to increase the performance of providing security by only providing cryptographic strategies on sensitive data in order to save memory and time.

KEYWORDS: Cloud Computing, Security Data Classification, Machine learning technique, Cryptanalytic Strategies

I. INTRODUCTION

In today's era Cloud Computing is the emerging virtual distributed environment that uses the ideas of storing, processing, sharing, connectivity and virtualization. Internet cloud facilitates large pool of communication between resources, storage media and sharing media that helps to supply on-demand services with minimal management effort. It employs parallel processing, distributed processing, grid computing and distributed database to enhance processing, in virtualization technology and the Internet broadband technology based on network [1].Innovators can focus on innovation rather than logistics of finding and managing resources that enable the innovation using cloud. Costs are driven down by delivering appropriate resources only at the time resources are required. The end users gain in complying with the ideas of isolation, elasticity, security and distribution. [2]
In cloud domain security issues are the foremost difficult problem and hence, the most key obstacle for aggrandize of IT-based companies that provide users on-demand services. These security issues can be visualizing at application phase, network phase authentication phase, information storage and at virtualization phase. These threats or challenges are still an obstacle within the complete success path of cloud computing. One reason is that consumers and plenty of organization keep their information on cloud database, so the main focus is that the user's data should be safe, and the vital information shouldn't get drift and tampered when travelling from one place to a different across the network. Thus, it is essential that Integrity, Confidentiality, and Availability of user information should to be ensured. Another reason is that, unauthenticated user tries to access the authenticated user's data.

Applying cryptographic algorithms in cloud servers to solve these threats is to done. Though when a user is revoked, using a single cryptography algorithm is not enough to assure the security levels i.e. confidentiality of data and managing the access control methods in cloud computing environment. For data security these techniques are applied

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 7, Issue 12, December 2018

on encryption and decryption. Encrypting complete data can turn to be very expensive in terms of memory as well as in terms of time and cost. So, to solve this problem it would be better if separation of sensitive data is done and then apply encryption algorithms. This would address reliable results if we classify the data [3] according to its sensitivity level.

The data classification is a method in machine learning field of distinguishing the category of unclassified data sample set with the help of classifiers. The classifier is constructed by constructing a training set of familiar data samples. To create an appropriate classifier, a large range of justified training data samples or data sets are required. This point of advancement invites a new paradigm of services where data classification is provided by servers in a cloud to its various clients or users. Specifically, the server will process the data automatically and hence, categorise the client's data samples present on remote private servers. However third-party-servers can access the private data.
Furthermore, any important data or training data set specifications may not be disclosed by the servers even if it provides the classification services to its client. Thus, a mechanism that ensures the privacy of the server's training set and client data samples is required. Hence, in this paper cryptographic algorithms are taken for encrypting and decrypting the data on the basis of sensitive data.

II. RELATED WORK

The great benefit for the users in cloud is accessibility, availability, confidentiality of data by internet. Also the data is stored at different servers so that the user can access data from anywhere and at anytime without bringing the physical storage devices. Shared data can be accessed by clients thereby enhancing the collaborative efforts. As there is no need to buy expensive hardware's, accessing through cloud storage can be economical. Moreover cloud facilitates recoveries from backup, disaster and archival. Despite these benefits of cloud there exist some major limitations. As in public cloud the information is present on remote servers and is not in control of legal users. Due to this the data can be accessed globally by unauthorised users and makes cloud computing model unsecure.
In consumer's perspective cloud computing issues was discussed in detail [4]. The main security issues discussed are: securing the information and secrecy protection. To protect data different techniques are suggested. The approach used was not practical and also rely on different methods for model execution. To secure the information mechanism in cloud, RSA techniques and AES techniques [5] was used in which guarantee the security of the users' information present on the cloud servers. The matter of fine-graininess, information confidentiality and quantifiability in cloud was discussed [6] .The different techniques like KP-ABE, re-encryption techniques (proxy and lazy) was used in the paper. Cryptography and steganography techniques were used for storing the data and information by author in [7]. The model has represented a 3 step data security model to secure the users' data. The steps involves: using cryptography algorithms with RSA, encrypting the data with stenography technique and in finally the accessing the data by decrypting via RSA algorithm. In this paper the work is mostly with sensitive data RSA is efficient to handle this limited given sensitive information.
A data classification model that result in minimum overhead and processing time in this paper. The different security mechanism with varying key length that provides data confidentiality was discussed. The model was evaluated with different encryption algorithms that show better result in terms of reliability and efficiency. The drawback of the paper was that the automatic classification of the data was needed. Moreover for higher degree of security and confidentially more secured cryptographic algorithms like RSA cryptography was required. The basic data i.e. non sensitive data can be done by AES cryptographic strategy as the data in this context is huge.
A Confidentiality based data classification model for cloud computing was suggested in [3]. In this paper modified K-NN (K Nearest Neighbour) technique was used for the classification of data. The main focus was to decide what kind of data need to be secured and which data should be kept as public. MK-NN was used for classification of data. The data was categorised into sensitive data and non-sensitive data. The confidential data was secured using RSA algorithm and the basic data was accumulated on the cloud servers. Drawback of the paper was that only RSA algorithm was used for sensitive data as it is limited. Moreover automatic recognition for the categories (basic, confidential, highly confidential) by MK-NN model was yet to be implemented.

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 7, Issue 12, December 2018

Most of these techniques used above were used for encrypting the data without considering its sensitivity level. Encrypting the whole data would be very expensive to its user's so to reduce the cost, first categorize the data into different classes (highly confidential, confidential or basic) and then apply the encrypting techniques only to most sensitive data. This would result in saving the encryption / decryption time and also become economical for its users.

III. PROPOSED WORK

Data classification is task of identifying data sets with respect to its data value. These values are based on usage consumption of data by its users and restrictions on access control methods [8]. KNN technique is used in machine learning, artificial intelligence that helps to categorise the class of unclassified data with the help of build classifier. It is constructed by employing a training set of familiar data samples.

In this proposed work, the modified algorithm of KNN i.e. MKNN Learning Technique is used to enhance the execution of existing KNN technique [9].This is costly, lazy, requires full training data plus depends on the value of k and has the issue of dimensionality because of the distance.

The main idea of the modified method is assigning the class label of the data according to K validated data points of the train set. Here the validity of all data samples in the train set is computed. Then, a weighted KNN is performed on any test samples. MKNN Pseudo-code:

```
Output_label := MKNN ( trained_set , test_sample )
Begin
 for  i := 1  to  trained_size
 validity(i) := Compute Validity of  i-th sample;
End  for;
 Output_label:=Weighted_KNN(Validity,test_sample);
 Return Output_label ;
End.
```

How modified model algorithm works:

Every sample in train set must be validated at the first step in MKNN algorithm. The validity of each point is computed according to its neighbours. The process of validation for all train samples is performed once. The validity of each train sample was assigned and it analysis as more information about the points

To validate a sample point in the train set, the H nearest neighbours of the point is considered. Among the H nearest neighbours of a train sample x, validity(x) counts the number of points with the same label to the label of x. The formula which is proposed to compute the validity of every point in train set is [10].

.

$$Validity(x) = \frac{1}{H}\sum_{i=1}^{H} S(lbl(x), lbl(N_i(x)))$$

Where H is the number of considered neighbours and lbl(x) returns the true class label of the sample x. also, Ni(x) stands for the i th nearest neighbour of the point x. The function S takes into account the similarity between the point x and the ith nearest neighbour.

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 7, Issue 12, December 2018

$$S(a,b)=\begin{cases}1 & a=b\\ 0 & a\neq b\end{cases}$$

Weighted KNN is one of the variations of KNN method which uses the K nearest neighbours, regardless of their classes, but then uses weighted votes from each sample rather than a simple majority or plurality voting rule.

Each of the K samples is given a weighted vote that is usually equal to some decreasing function of its distance from the unknown sample. For example, the vote might set be equal to 1/(de+1), where de is Euclidian distance. These weighted votes are then summed for each class, and the class with the largest total vote is chosen. This distance weighted KNN technique is very similar to the window technique for estimating density functions. For example, using a weighted of 1/(de+1) is equivalent to the window technique with a window function of 1/(de+1) if K is chosen equal to the total number of training samples [11]

In the MKNN method, first the weight of each neighbour is computed using the 1/(de+0.5). Then, the validity of that training sample is multiplied on its raw weight which is based on the Euclidian distance. In the MKNN method, the weight of each neighbour sample is derived according to [12]

$$W(i)=Validity(i)\times\frac{1}{d_e+0.5}$$

Here W(i) and Validity(i) are for the weight and the validity of the i^{th} nearest sample in the train set. This learning technique has the effect of giving greater importance to the reference samples that have more validity and closeness to the test sample. Hence, the decision is less affected by reference samples which are not very stable in the feature space in comparison with other samples.
Other way, the multiplication of the validity measure on distance based measure can overcome the weakness of any distance based weights which have many problems in the case of outliers. Based upon distance the proposed MKNN algorithm is significantly stronger than the traditional KNN method.

Cloud simulation environment:

The cloud simulation environment based on this proposed work is as shown in fig-1, is used for the proposed model that solves the problem of confidentiality and classification of data. Firstly we run the simulation then data centres' are initiated. The virtual machine manager (VMM) is used to handle the VMs and for allocating VMs to cloudlets i.e., the cloud task. Authentication process is performed so that only authenticated users can access the data. Then classification KNN is enhanced with modified KNN learning technique. This will improve the prediction capabilities and accuracy of existing KNN.

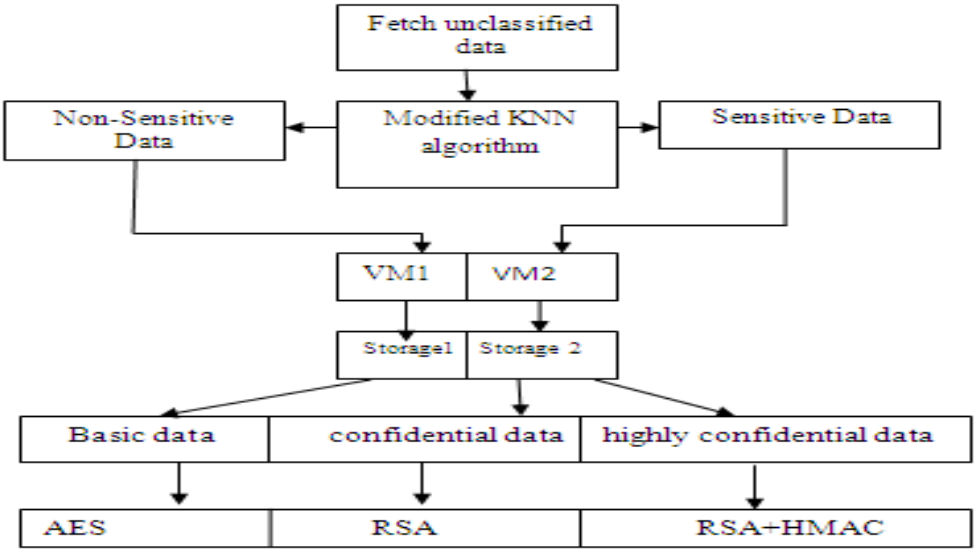

Figure 1: Proposed Data Classification System with cryptographic strategies

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 7, Issue 12, December 2018

IV. RESULTS AND DISCUSSIONS

The proposed method is evaluated on five sample standard data sets, namely Wine, Isodata and Monk's problems. None of the databases had missing values, as well as they use continuous attributes. These data sets which are obtained from UCI repository [13] are described as follows

TABLE I. COMPARISON OF RECOGNITION RATE BETWEEN THE MKNN AND KNN ALGORITHM (%)

		Monk 1	Monk 2	Monk 3	Isodata	Wine
K=3	KNN	84.49	69.21	89.12	82.74	80.89
	MKNN	**87.81**	**77.66**	**90.58**	**83.52**	**83.95**
K=5	KNN	84.26	69.91	89.35	82.90	83.79
	MKNN	**87.81**	**78.01**	**90.66**	**83.32**	**85.76**
K=7	KNN	79.86	65.74	88.66	80.50	80.13
	MKNN	**86.65**	**77.16**	**91.28**	**83.14**	**82.54**

The experiments show the MKNN method significantly outperforms the KNN method, with using different choices of value K, over different datasets. More performance is because of that the classification is based on validated neighbours which have more information in comparison with simple class labels. The value of parameter H is determined equal to a fraction of the number of train data which is empirically set to 10% of the train size. In addition, since computing the validity measure is executed only once in training phase of the algorithm, computationally, the MKNN method can be applied with the nigh same burden in comparison with the weighted KNN algorithm.

Now to secure highly confidential data we combine RSA encryption technique with HMAC. The HMAC (hashed message authentication code) is a cryptographic checksum. It is stored at local machine as well as appended with the cipher text and sent to the cloud. The HMAC+ cipher text must be matched with the HMAC checksum i.e, already stored at the local machine of user. The hacker finds it difficult to guess the HMAC checksum while hacking the data. Thus, results in increasing the security of user's HC data by using RSA encryption technique.
HMAC is calculated by following formula:
HMAC(Key,msg)=HMAC(KeyXORout)||H((Key XORin)||msg)) where Key is the original key, m is the message that needs to be authenticated.

V. CONCLUSIONS AND FUTURE SCOPE

This paper aims at achieving data confidentiality, data access control management and provides an automatic classification of data in cloud [14]. On the contrary, the existing system was lacking in providing automatic classification of data. In this paper the functionality and performance of modified KNN Learning technique. The data will be automatically classified by the machine on the basis of data security parameters. Also to emphasis more focus on highly confidential data HMAC function has been appended with the existing RSA encryption algorithm. The proposed system proves to be more accurate and economical. And also saves the user time for encrypting/decrypting different classes of data (basic, confidential and highly confidential).

REFERENCES

[1]Application architecture for Cloud Computing,IBM,WHITE PAPER,Nov 2009.
[2]F. F. Moghaddam, M. Vala, M. Ahmadi, T. Khodadadi, and K. Madadipouya, "A reliable data protection model based on re-encryption concepts in cloud environments," Proc. - 2015 6th IEEE Control Syst. Grad. Res. Colloquium, ICSGRC 2015, pp. 11–16, 2016.

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 7, Issue 12, December 2018

[3]M. A. Zardari, L. T. Jung, and N. Zakaria, "K-NN classifier for data confidentiality in cloud computing," 2014 Int. Conf. Comput. Inf. Sci. ICCOINS 2014 - A Conf. World Eng. Sci. Technol. Congr. ESTCON 2014 - Proc., 2014.
[4]D. Chen and H. Zhao, "Data Security and Privacy Protection Issues in Cloud Computing," 2012 Int. Conf. Comput. Sci. Electron. Eng., no. 973, pp. 647–651, 2012.

[5]N. Surv, B. Wanve, R. Kamble, S. Patil, and J. Katti, "Framework for client side AES encryption technique in cloud computing," Souvenir 2015 IEEE Int. Adv. Comput. Conf. IACC 2015, pp. 525–528, 2015.
[6]S. Yu, C. Wang, K. Ren, and W. Lou, "Achieving secure,scalable ,and fine-grained data access control in cloud computing.pdf," Ieee Infocom, pp. 1–9, 2010.
[7]V. K. Pant, J. Prakash, and A. Asthana, "Three step data security model for cloud computing based on RSA and steganography," 2015 Int. Conf. Green Comput. Internet Things, pp. 490–494, 2015.
[8]L. Tawalbeh, N. S. Darwazeh, R. S. Al-Qassas, and F. AlDosari, "A secure cloud computing model based on data classification," Procedia Comput. Sci., vol. 52, no. 1, pp. 1153–1158, 2015.
[9] Fix, E., Hodges, J.L. Discriminatory analysis, nonparametric discrimination: Consistency properties. Technical Report 4, USAF School of Aviation Medicine, Randolph Field, Texas, 1951.
[10] L. I. Kuncheva, Combining Pattern Classifiers, Methods and Algorithms, New York: Wiley, 2005
[11] Cover, T.M., Hart, P.E. Nearest neighbor pattern classification. IEEE Trans. Inform. Theory, IT-13(1):21–27, 1967.
[12] E. Gose, R. Johnsonbaugh and S. Jost, Pattern Recognition and Image Analysis, Prentice Hall, Inc., Upper Saddle River, NJ 07458, 1996.
[13]Blake,C.L.Merz, C.J. UCI Repository of machine learning databases,http://www.ics.uci.edu/~mlearn/ ML Repository.html, (1998).
[14]R. Shaikh and M. Sasikumar, "Data classification for achieving security in cloud computing," Procedia Comput. Sci., vol. 45, no. C, pp. 493–498, 2015.