International Journal of Pure and Applied Mathematics
Volume 119 No. 14 2018, 1709-1717
ISSN: 1314-3395 (on-line version)
url: http://www.ijpam.eu
Special Issue

ijpam.eu

INTRUSION DETECTION IN CLOUD COMPUTING USING ARTIFICIAL NEURAL NETWORK

Prasanta Kumar Bal[1], Sateesh Kumar Pradhan[2] , Sudhir Kumar Mohapatra[3]
[1]Research Scholar, [2]Professor, PG Dept of CS, Utkal University, Odisha, India
[3]Associate Professor, Dept of CSE,IT & SE, AASTU, Addis Ababa, Ethiopia-16417

ABSTRACT:

Currently cloud computing provides computing and data storage services through the Internet. Increasing in amount of cloud users raises the privacy and Security concerns. Data security became the major concern as the user's data managed with a third party. In this paper, an efficient technique is developed in order to facilitate information security while updating the data within the cloud. The technique utilizes two techniques such as Fuzzy c means (FCM) clustering and Artificial Neural Network technique (ANN). The FCM perform grouping of the existing data into clusters and the ANN train itself with the data simultaneously. Initially, the new data is passed to FCM, where it is clustered. The data which cannot be clustered are forwarded to ANN. The anomaly data is predicted at the ANN, which is then queued. The proposed technique is implemented in the working platform of JAVA and the performance is analyzed.

INTRODUCTION:

Cloud computing, a new internet-based technology, has been widely envisioned as the most promising technology of IT enterprise [12]. Thus the word cloud computing can be defined as utilizing the internet to provide technology enabled services to the people and organizations. Cloud computing is independent computing .It is totally different from grid and utility computing [7]. Cloud computing is based on five attributes: multi-tenancy (shared resources), massive scalability, elasticity, pay as you go, and self-provisioning of resources [6]. Cloud computing provides a facility that enable large scale controlled sharing and interoperation among resources that are dispersedly owned and managed [7].

Cloud computing can be divided into three levels: the infrastructure layer, platform services layer, application layer software [8]. Cloud computing has been a paradigm shift in the information technology domain [11], that delivers highly scalable distributed computing platforms in which computational resources are offered as a service [10]. Cloud Computing refers to both the applications delivered as services over the Internet and the hardware and systems software in the data centers that provide those services [9]. Specifically, the cloud servers managed by cloud providers are not fully trusted by users while the data files stored in the cloud may be sensitive and confidential, such as business plans.

An Intrusion Detection System (IDS) is software and/or hardware, which is designed for identifying the undesirable efforts for enhancing the computer security systems [13]. Especially, the wireless sensor devices has given rise to a wider range of amazing applications in various walks of our life that involve environment and habitat monitoring, healthcare applications and many more. But, simultaneously, the sensor nodes have produced the same number of threats caused by attackers, whose intention is to achieve access to the network and the data transferred inside it. Till now, numerous classical security methodologies exist for the purpose of avoiding these intrusions [14].The intrusion detection systems fall into two important categories. One category is for analyzing the network traffic and the other is to analyze the operating system audit trails. These systems use either the rule-based misuse detection or anomaly detection naturally [15] and their power relies on the ability of the security personnel developing them to a larger extent.

With the coming of Internet age, network security has become the key foundation to web applications such as online retail sales, online auctions, etc. Intrusion detection attempts to detect computer attacks by examining various data records observed in processes on the network. Detection precision and detection stability are two key indicators to evaluate intrusion detection systems (IDS). In order to enhance the detection precision and detection stability, many researchers have been done and in the

early stage, the research focus lies in using rule based expert systems and statistical approaches. But when encountering larger datasets, the results of rule based expert systems and statistical approaches become worse [16]

1. RELATED WORK:

Several techniques are available in the literature for detecting the intrusion behavior. In recent times, intrusion detection has received a lot of interest among the researchers since it is widely applied for preserving the security within cloud.

G. Gowrisona et al. [1] designed an intrusion detection system to classify by the incorporation of enhanced rules as learnt from the network behavior with less computational complexity of O (n). The method demonstrates the achievements of promising classification rate.

Shingo Mabu et al [2] described a fuzzy class association rule mining method based on genetic network programming (GNP) for detecting network intrusions. GNP is an evolutionary optimization technique, which uses directed graph structures instead of strings in genetic algorithm or trees in genetic programming, which leads to enhancing the representation ability with compact programs derived from the reusability of nodes in a graph structure. By combining fuzzy set theory with GNP, the method can deal with the mixed database that contains both discrete and continuous attributes and also extract many important class association rules that contribute to enhancing detection ability. Therefore, the method can be flexibly applied to both misuse and anomaly detection in network-intrusion-detection problems.

M. Bahrololum et al. [3] proposed an approach to design the system using a hybrid of misuse and anomaly detection for training of normal and attack packets respectively. The utilized method for attack training was the combination of unsupervised and supervised Neural Network (NN) for Intrusion Detection System. By the unsupervised NN based on Self Organizing Map (SOM), attacks were classified into smaller categories considering their similar features, and then unsupervised NN based on Back propagation were used for clustering. By misuse approach known packets were identified fast and unknown attacks were able to detect by this method.

K.S. Anil Kumar and V. Nanda Mohan [4] proposed a combination of three techniques comprising two machine-learning paradigms. K-Means Clustering, Fuzzy Logics and Neural Network techniques were deployed to configure an effective intrusion detection system. Out of the several problems in the traditional techniques of Intrusion Detection Systems, the presence of high rate of false alerts caused unnecessary interference of human analyst. The human analysts in turn performed an intensive analysis repeatedly to distinguish the nature of such alerts and initiate sufficient actions. The approach proposed revealed the advantage of converging K-Means-Fuzzy-Neuro techniques to eliminate the preventable interference of human analyst in such occasions.

Latifur Khan et al. [5] have proposed a method that is a scalable solution for detecting network based anomalies. The proposed method presented a study for enhancing the training time of SVM, specifically when dealing with large data sets, using hierarchical clustering analysis. They utilized the Dynamically Growing Self-Organizing Tree (DGSOT) algorithm for clustering because it has proved to overcome the drawbacks of traditional hierarchical clustering algorithms (e.g., hierarchical agglomerative clustering). Clustering analysis was helpful to find the boundary points, which are the most qualified data points to train SVM, between two classes.

2. INTRUSION DETECTION IN CLOUD COMPUTING:

Cloud computing is a ground breaking computing model providing resources and applications as something over the Internet for satisfying the computing demand of the users. The environment of the Cloud networking is changed according to the people and their environments. Various numbers of clients are connected with number of servers can provide different types of resources. The resources can be accessed by the clients through the servers. The cloud architecture is given as by the following figure 1,

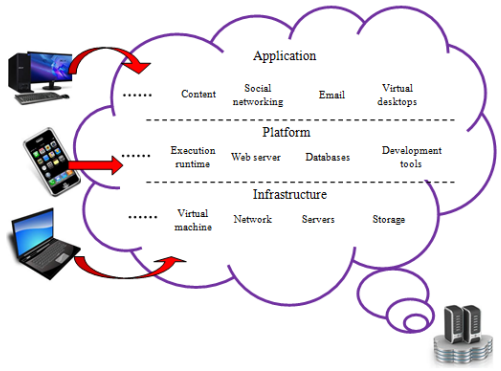

Figure 1: Cloud Architecture

One of the major issues in cloud is to detect the intrusion data in order to ensure the information security. Thus the proposed technique is emerged so as to detect and discard the intrusion data. The proposed technique is explained in detail in the following sections.

2.1 Proposed Intrusion data Prediction Method:

The proposed intrusion detection method to realize the dividing of normal and abnormality behaviors of the data . This is checked while making any upload and update of data. The update of data refers to the process of adding newer data. While the term 'upload' refers to the process of adding newer dataset. Data uploading is proceeded by simply reading and executing into the cloud.

The data updating is checked here for the normal/anomaly data. So that, the proposed technique to detect the intrusion data is performed by means of the Fuzzy c-means clustering and the ANN. The data within the cloud is initially clustered and trained with the help of FCM and ANN techniques respectively. Then, while updating newer data, it is checked for any possibility to group along with the existing clusters. If the clustering is not made properly, it is checked with the ANN architecture.

The process taking place within the FCM and ANN are given in the below sections.

2.2 Fuzzy c-means:

Fuzzy c-means (FCM) clustering algorithm can flexibly divide the sample data which has been used into intrusion detection. In our proposed technique, the intrusion data is separated after clustering. When applying FCM to intrusion detection to realize dividing normal and abnormality behaviors, it will have more evident effect for the larger number of clustering samples.

In FCM, the data points can belong to more than one cluster and the relationship between the data points are determined by means of the membership grade. The membership grade gives the degree to which the data point is related to various clusters.

The proposed architecture diagram is shown by the below figure 2.

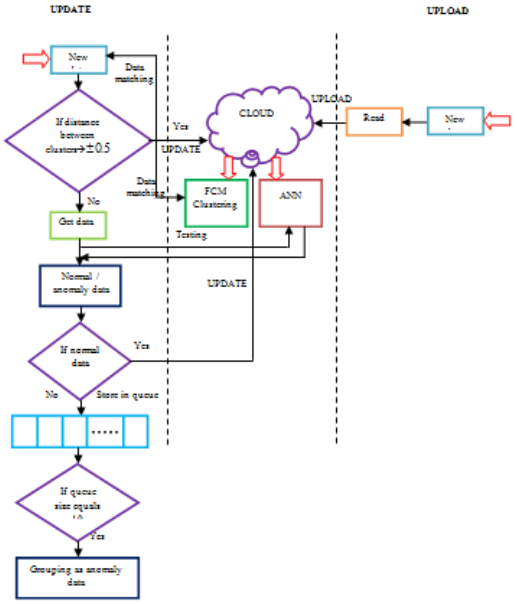

Figure 2: Proposed Architecture

During Fuzzy c-means clustering, the input data will be grouped into number of clusters randomly and the centroids will be generated for the clusters. At the each iteration, the clusters are updated based on the membership grade of the data points and the new centroid is determined correspondingly. Moreover, how the clustering with fuzzy c means algorithm is made for a set of input samples is given below.

Let us considering the input sample be,

$$Y_m\,(m=1,2,...,p) \quad (1)$$

The input sample is to be divided into 's' number of clusters. The clustering cannot be exactingly but it will be made by means of the grouping with respective to the grade of membership function.

The membership matrix is of as follows:

$$R=[r_{km}]_{s\times p} \quad (2)$$

Where,

r_{km} - Membership grade of $k(k=1,2,...,p)$ and $m(m=1,2,...,s)$

Also, the value within the membership matrix must meet the following criteria which is given as,

$$\begin{cases} \sum_{k=1}^{s} r_{km}=1, \ 1\le m\le P \\ 0\le r_{km}\le 1, \ 1\le k\le sP, 1\le m\le P \\ 0<\sum_{k=1}^{s} r_{km}<P, \ 1\le k\le s \end{cases} \quad (3)$$

The clustering will be made by means of calculating the minimum distance between the data units. So, the objective function of FCM is made with the help of integrating the minimal distance along with the membership matrix. The objective function is given as below,

$$\min L(R,T)=\sum_{k=1}^{s}\sum_{m=1}^{p} r_{km}^{g}\, e_{km}^{2} \quad (4)$$

Where,

$T=(T_1,T_2,...,T_s)$ is the center of clustering

e_{km} - Euclidean distance

The Euclidean distance can be given through the following representation as,

$$e_{km}=\|Y_m-T_k\| \quad (5)$$

Here, the clustering into normal data is made when the Euclidean distance is of, ± 0.5.

Furthermore, g is the weighted index to control the fuzzy degree of membership matrix, (R, g). The value of g must be greater than one ($g>1$) and the greater values represent the fuzzier form of membership matrix. For the value equal to one ($g=1$), fuzzy clustering will be degenerated to hard c-means clustering.

Using Lagrange multiplier method, the parameters required for the computation is minimized as follows,

$$r_{km}=\left(\sum_{l=1}^{s}\left(\frac{\|Y_k-T_m\|}{\|Y_l-T_m\|}\right)^{\frac{2}{g-1}}\right)^{-1} \quad (6)$$

$$T_k=\frac{\sum_{m=1}^{s}(r_{km})^{g}Y_g}{\sum_{m=1}^{s}(r_{km})^{g}} \quad (7)$$

Intended for FCM algorithm, the 's' and 'g' are firstly fixed and an initial membership matrix can be set arbitrarily. According to equation (6) and equation (7), the classification and clustering centers are adjusted by time.

The algorithm will be regarded as to be converged before the change of clustering centers between two adjoining iterations is very small or perhaps the change value of objective function is smaller when compared to a threshold value. In that case, the membership value of each sort of clustering center and each group with each sort can be received. With the max membership principle in fuzzy sets, the type of each sample point can be confirmed. After that the label of fuzzy clustering can be finished.

While updating the cloud with newer data, the data is checked with all the clusters on where it should be added. The cluster with distance within the range of ± 0.5, is selected. If no clusters were within that particular range, the data is predicted as anomaly data. Further, the data is checked with NN.

2.3 Artificial Neural Network:

An artificial neural network (ANN) is a way of reasoning that is influenced by principles studied in neurons of the central nervous system of life. The Neural network is

composed of three layers namely, input layer, hidden layer and output layer. Each layer has some artificial neurons (nodes), a weight matrix, and an output vector, and each neuron has a bias.

A layer of neurons that receives inputs directly from beyond the network is called input layer, a level that produces the outcome of network is referred to as output layer and layers that are between the output and input layers are called hidden layers. Multiple layers of neurons with nonlinear transfer functions allow the network to learn linear and nonlinear associations between input and output parameters.

In feed-forward systems, individual inputs are multiplied by weights and then, the weighted value is fed to the summing junction and is summed with the bias of the neuron. The back propagation neural network is proposed here. The back propagation NN is same as the feed forward NN, but the only difference is that the error is back propagated.

The datas within the cloud are initially trained by the neural network. Let the data be, $a_m = data_1, data_2, ..., data_m$. The data are passed to the hidden layer. The process within the hidden layer is shown below:

$$H_l = \left(\sum_{m=1}^{M} Weight_{lm} \times a_m \right) + Bias_l \quad (8)$$

Where,

H_l - Hidden layer at l^{th} node

$Weight_{lm}$ - Interconnection weight of input vector a_m at l^{th} node

$Bias_l$ -bias at l^{th} node

M -Number of elements

The ANN architecture is shown in figure 3,

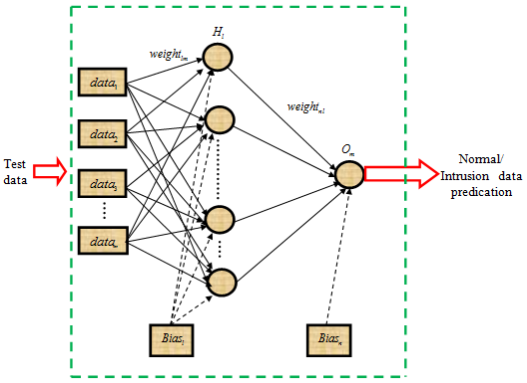

Figure 3: ANN Architecture

The output is generated at the hidden layer by taking the transfer function of input. Most commonly tan sigmoid function is taken as this activation function. The tan sigmoid function is given as,

$$\phi(H_l) = \frac{e^{H_l} - e^{-H_l}}{e^{H_l} + e^{-H_l}} \quad (9)$$

Eq. (9) is the output at l^{th} node, which is forwarded to the next layer. The process takes place for all the hidden neurons and are forwarded to the output layer. At the output layer, the weighted sum of the hidden layer output (i.e. previous layer) and the bias function is passed through the activation function, which is given as follows:

$$O_m = \phi\left[\left(\sum_{l=1}^{L} Weight_{nl} \times \phi(H_l) \right) + Bias_n \right] \quad (10)$$

Where,

$Weight_{nl}$ - Interconnection weight of hidden vector

L - number of elements

At last, the error between the target and the estimated outcomeis determined. This can be achieved by means of computing the MSE between the target and the predicted outcome. The error determination is given as,

$$MSE = \frac{1}{Q} \sum_{q=1}^{Q} \left(O_{t\arg et,q} - O_{predicted,q} \right)^2 \quad (11)$$

Then, the determined error value is back propagated to the hidden layer. Further, the interconnection weight is adjusted in such a way that the error value must be reduced. The process continues until the error value equals zero (or) the value becomes lesser than 0.0001. This is the training process and the testing process is carried out for the newer data.

As the datas within the cloud is trained within the Neural Network, the training weight is generated and is set as the testing weight for testing the anomaly data. The anomaly data is checked with the trained data. If the test score is greater than 5, the data is updated in the cloud. If not, the data is added in queue. The data addition into the queue is again checked for the possibility of grouping when the queue is filled with ten data entry. This datas are considered as the intrusion data. If the data to be updated is detected as intrusion data, it will be neglected or stored separately for future analysis.

3. IMPLEMENTATION

An implementation has been carried out using Java programming language to compare our proposed algorithms efficiency and accuracy for Intrusion Detection System Which has been supported by Artificial Neural Network Algorithm. The testing environment consists of four different numbers of updated data.

3.1 Performance evaluation

The various parameter used to describe the performance of our proposed algorithm is discussed as follows

False Acceptance can be defined as number of imposter person being authenticated as genuine because the criteria of reference threshold is fulfilled and the total number of imposter person is lying in the range of genuine person. It is defined in

$$FAR=1-\frac{\text{Wrongly accepted individual}}{\text{Total number of matching}} \quad (1)$$

The negative predictive value is defined as:

$$FPR=1-\frac{\text{number of true negatives}}{\text{number of true negatives+number of false negatives}} \quad (2)$$

where a "true negative" represent that the test makes a negative prediction, and the subject has a negative result under the gold standard, and a "false negative" represent that the test makes a negative prediction, and the subject has a positive result under the gold standard.

Accuracy (AC): can be defined as the proportion of the total number of the correct predictions to the actual data set size". As given in below equation

$$\text{Accuracy}=\frac{TN+TP}{TN+TP+FN+FP} \quad (3)$$

The effectiveness our proposed algorithm is checked using Intrusion Detection Systems Here we find the most basic factors such as true positive, true negative, false positive, false negative values they are listed in the below Table1

True Negatives (TN) as well as True Positives (TP) corresponds to correctly classified instances, that is, events that are rightly labeled as normal and attacks, respectively. Alternatively, False Positives (FP) refer to normal events being labeled as attacks while False Negatives (FN) are attack events incorrectly predicted as normal events. if the number of data transmitted is five then it's True Positives value is 3, True Negatives is 1, False Positives is 0, False Negatives is 1 but if the number of data transmitted is ten then it's True Positives value is 7, True Negatives is 1, False Positives is 1, False Negatives is 1 like that if the number of data transmitted is fifteen then it's True Positives value is 11, True Negatives is 2, False Positives is 1, False Negatives is 1 if the number of data transmitted is twenty then it's True Positives value is 15, True Negatives is 2, False Positives is 1, False Negatives is 2

Number of data updated	TP	TN	FP	FN
5	3	1	0	1
10	7	1	1	1
15	11	2	1	1
20	15	2	1	2

Table1: Intrusion Detection Systems TP, TN, FP, FN values

The performance of our proposed method in terms of accuracy, False Prediction Ratio and False Acceptance Ratio were calculated as follows if the number of data transmitted is 5,10,15,20 then its accuracy calculated will be 0.8,0.8,0.86667,0.85. Here False Prediction Ratio for the various number of updated data is 0.5,0.5,0.4,0.5 while False Acceptance Ratio for the various number of updated data is 0.25,0.125,0.1,0.2 respectively.

Accuracy	False Prediction Ratio	False Acceptance Ratio
0.8	0.5	0.25
0.8	0.5	0.125
0.86667	0.4	0.1
0.85	0.5	0.2

Table2: Accuracy, False Prediction Ratio, False Acceptance Ratio of updated data

The accuracy of our proposed method is shown in below figure4 in that accuracy values are high indicates its performance using our proposed method is good.

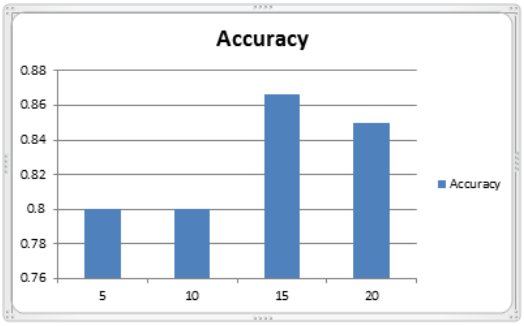

Figure4 proposed method accuracy

The False Prediction Ratio of our proposed method is shown in below figure5 here for all the updated data False Prediction Ratio obtained is very high it indicates that our proposed method has better prediction.

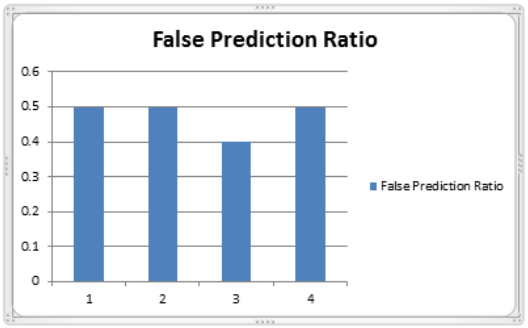

Figure 5 proposed method False Prediction Ratio

The False Acceptance Ratio of our proposed method is shown in below figure6 here for all the updated data False Acceptance Ratio obtained is very low it indicates our proposed method has better data transmission acceptance rate.

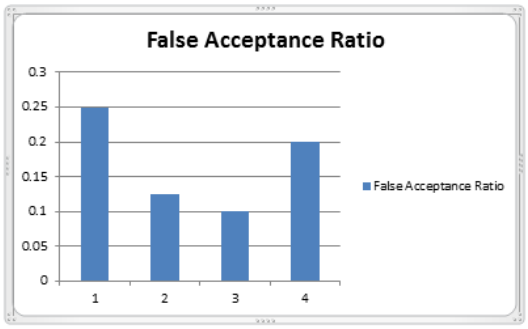

Figure6 proposed method False Acceptance Ratio

The data updated using our proposed method has better Accuracy, False Prediction Ratio, False Acceptance Ratio. it is prevented from various types of attacks and data transmitted in this were highly secure.

4. CONCLUSION:

Here, in order to facilitate the information security, the proposed technique is developed. The performance of the proposed is analyzed to show the efficiency of the technique. More over, the performance results were developed for the proposed as well as the existing technique to show the efficiency of the proposed technique over the existing techniques. Moreover, the analysis from the results shows the better prediction of intrusion data than the existing ones.

REFERENCES:

[1] G. Gowrisona, K. Ramarb, K. Muneeswaranc, T. Revathic, " Minimal complexity attack classification intrusion detection system", Applied Soft Computing, vol 13, pp: 921–927, 2013.

[2] Shingo Mabu, Nannan Lu, Kaoru Shimada,Kotaro Hirasawa, " An Intrusion-Detection Model Based on Fuzzy Class-Association-Rule Mining Using Genetic Network Programming", IEEE Transactions On Systems, Man, And Cybernetics—Part C: Applications And Reviews, VOL. 41, NO. 1, PP: 130-139, 2011

[3] M. Bahrololum, E. Salahi and M. Khaleghi "Anomaly intrusion detection design using hybrid of unsupervised and supervised neural networks", International Journal of Computer Networks & Communications, Vol.1, No.2, 2009

[4] K.S. Anil Kumar and Dr. V. NandaMohan, " Novel Anomaly Intrusion Detection Using Neuro-Fuzzy Inference System ", IJCSNS International Journal 6 of Computer Science and Network Security, vol.8, no.8, pp.6-11 , August 2008.

[5] Latifur Khan, Mamoun Awad, Bhavani Thuraisingham, "A new intrusion detection system using support vector machines and hierarchical clustering", The International Journal on Very Large Data Bases, Vol. 16, no. 4, October 2007.

[6] EmanM.Mohamed, Hatem S.Abdelkader and SherifEI-Etriby, "Enhanced Data Security Model for Cloud Computing", In proceeding of IEEE International Conference on INFOrmatics and Systems, pp. 12-17, May 2012.

[7] Engr: Farhan Bashir Shaikh and Sajjad Haider, "Security Threats in Cloud Computing", In proceeding of IEEE International Conference on Internet Technology and Secured Transactions, pp. 214-219, Dec 2011.

[8] Zhang Xin, Lai Song-qing and Liu Nai-wen, "Research on Cloud Computing Data Security Model Based on Multi-dimension", In proceeding of IEEE International Symposium on Information Technology in Medicine and Education, Vol. 2, pp. 897-900, Aug 2012.

[9] Xu Wang, Beizhan Wang and Jing Huang, "Cloud computing and its key techniques", In proceeding of IEEE International Conference on Computer Science on Automation Engineering, Vol. 2, pp. 404-410, Jun 2011.

[10] Mohemed Almorsy, John Grundy and Amani S. Ibrahim, "Collaboration-Based Cloud Computing Security Management Framework", In proceeding of IEEE International Conference on Cloud Computing, pp. 364-371, July 2011.

[11] Murat Kantarcioglu, Alain Bensoussan and SingRu(Celine) Hoe, "Impact of Security Risks on Cloud Computing Adoption", In proceeding of IEEE International Conference on Communication, Control and Computing, pp. 670-674, Sep 2011.

[12] Shuai Han and Jianchuan Xing, "Ensuring Data Storage Security Through A Novel Third Party Auditor Scheme in Cloud Computing", In proceeding of IEEE International Conference on Cloud Computing and Intelligence Systems, pp. 264-268, Sep 2011

[13] Disha Sharma," Fuzzy Clustering as an Intrusion Detection Technique", International Journal of Computer Science & Communication Networks, Vol.1, No.1, 2011

[14] Ioannis Krontiris, Zinaida Benenson, Thanassis Giannetsos, Felix C. Freiling and Tassos Dimitriou, "Cooperative Intrusion Detection in Wireless Sensor Networks", Lecture Notes in Computer Science, Vol. 5432, pp 263-278, 2009

[15] Peter Lichodzijewski, A. Nur Zincir-Heywood and Malcolm I. Heywood, "Host-Based Intrusion Detection Using Self-Organizing Maps", Fac. of Comput. Science

[16] Swati Ramteke, Rajesh Dongare, Komal Ramteke, "Intrusion Detection System for Cloud Network Using FC-ANN Algorithm," International Journal of Advanced Research in Computer and Communication Engineering, Vol. 2, 2013