# An Efficient Optimized Fuzzy Inference System based Intrusion Detection in Cloud Environment

**S. Immaculate Shyla, S. S. Sujatha**

***Abstract*****:** ***Security incidents namely, Denial of service (DoS), scanning, virus, malware code injection, worm and password cracking are becoming common in a cloud environment that affects the company and may produce an economic loss if not detected in time. These problems are handled by presenting an intrusion detection system (IDS) in the cloud. But, the existing cloud IDSs affect from low detection accuracy, high false detection rate and execution time. To tackle these issues, in this paper, gravitational search algorithm based fuzzy Inference system (GSA-FIS) is developed as intrusion detection. In this approach, fuzzy parameters are optimized using GSA. The proposed consist of two modules namely; Possibilistic Fuzzy C-Means (PFCM) algorithm based clustering, training based on GSA-FIS and testing process. Initially, the incoming data are pre-processed and clustered with the help of PFCM. PFCM is detecting the noise of fuzzy c-means clustering (FCM), conquer the coincident cluster problem of Possibilistic Fuzzy C-Means (PCM) and eradicate the row sum constraints of fuzzy Possibilistic c-means clustering (FPCM). After the clustering process, the clustered data are given to the optimized fuzzy Inference system (OFIS). Here, normal and abnormal data are identified by the Fuzzy score, while the training is done by the GSA through optimizing the entire fuzzy system. In this approach, four types of abnormal data are detected namely, probe, Remote to Local (R2L), User to Root (U2R), and DoS. Simulation results show that the performance of the proposed GSA-FIS based IDS outperforms that of the different scheme in terms of precision, recall and F-measure.***

***Keywords*** **:** ***Gravitational search algorithm, Possibilistic Fuzzy C-Means, cloud computing, intrusion detection system, R2L, U2L, DOS, probe, fuzzy Inference system.***

## I. INTRODUCTION

These days, cloud computing [1] renders information storage and computing administrations through the Internet. Cloud computing has speed, versatility, and flexibility, and so forth. By the CSP, cloud computing is an ordinary term for whatever surrenders passing on encouraged administrations over the Internet and managed. At remote areas, the Cloud administrations permit associations and people to use programming and hardware framework that is managed by outcasts. Information security gets the standard issue as the customer's information managed by an outsider [2]. The amount of assaults has grown comprehensively, and a couple of novel hacking gadgets and meddling procedures have shown up with the in all cases usage of PC frameworks. Applying IDS is one technique for dealing with suspicious activities inside a framework [3]. An IDS screens the activities of an oversaw area and chooses whether these exercises are pernicious (intruded) or bona fide (unique) set up on system genuineness, classification and the availability of information affirms [4]. The IDSs might be adjusted to do abuse detection or irregularity detection by and large [5]. All realized anomalous conduct is resolved and the framework is prepared to recognize it in abuse detection. It works by contrasting the arriving bundle and attributes of realized assault direct. If any new, not predefined assault arrives, the structure would remember it as a common bundle, instigating high FNR [6].To disregard high FNR, misuse based IDS must be retrained constantly, now and again initiating deferrals in the framework [7]. Along these lines, chose run of the mill lead irregularity recognition is shown [8], so any model dismissing that direct would be settled as system assault [5]. Irregularity location initiates high FPR, on the grounds that even novel run of the mill parcel, darken to the system, and would be perceived as an assault. Cloud computing moreover encounters different regular assaults, for example, IP caricaturizing, Address Resolution Protocol taunting, Routing Information Protocol assault, DNS hurting, Flooding, DoS, Distributed DoS [9]. To raise the gathering of Web and cloud administrations, CSPs ought to at first show trust and security to lessen the worries of a tremendous number of customers. A strong cloud biological system should be free from abuse, viciousness, swindling, hacking, diseases, lament, suggestive excitement, spam, and protection and copyright encroachment [10]. ID in cloud computing is an NP-Hard problem. Therefore, a lot of optimization and meta-heuristic methods are developed to solve the problem. Statistical-based IDS utilizes different statistical methods admitting principal component analysis, cluster and multivariate analysis, Bayesian analysis, and frequency and simple important analysis. Be that as it may, this sort of IDS requires accumulating enough information to assemble a complex mathematical model, which is impractical on account of complicated system traffic. To determine the impediments of above strategies, various information mining techniques have been initiated [11] in particular, support vector machine (SVM), artificial neural network (ANN), Fuzzy logic system (FLS), k-nearest neighbor (K-NN), deep learning and navie Bayes, etc.

**Revised Manuscript Received on October 30, 2019.**
***** Correspondence Author

**S.Immaculate Shyla***, Research Scholar, Registration Number: 17223152162007, Department of Computer Science, Manonmaniam Sundaranar University, Tirunelveli, India. Email: immaculatejudit@gmail.com

**Dr.S.S.Sujatha,** Department of Computer Applications, S.T.Hindu College, Nagercoil, India Email: sujaajai@gmail.com

Even though, this algorithm also has some drawback such as minimum classification precision, especially for low-frequent attacks, e.g., R2L, U2R, and weaker detection stability [12]. The main aim of this study is to detect the different type of attack data based on a combination of PFCM and OFIS. Our proposed method comprised of three stages namely, clustering, training, and testing. Initially, the dataset is divided into two subsets such as training and testing. Then, the trained data are given to the preprocessing stage. Then, the preprocessed data are given to the clustering process. For clustering, PFCM classifier is utilized. After the clustering process, the clustered data are given to the OFIS. In FIS, the whole fuzzy system is optimally selected with the help of GSA algorithm. Finally, in the testing process, the classifier is detected, the given data as normal or attack data. Then, the normal data are stored in the cloud.

Contributions of this proposed approach are described as follows:

- Four attacks such as a probe, U2R, R2L, and DOS are considered in this approach
- For clustering, PFCM algorithm is developed. This is used to reduce the complexity and time.
- For IDS, a fuzzy inference system (FIS) is developed. This FIS system is optimized by using gravitational search algorithm in this approach.
- This proposed approach is implemented in the platform of JAVA.
- From the implementation results, it is proved that the performance of this proposed approach is superior to that of existing work in terms of precision, recall, and F-measures

The paper is organized as follows: the related works presented in section 2 and proposed methodology is presented in section 3. The experimental results are analyzed in section 4 and the conclusion part is presented in section 5.

## II. RELATED WORKS

The researchers are more interested in Intrusion detection since it is normally maintaining security over the network at current days. Here, they represented some of the intrusion detection techniques.

**Table 1: Literature Survey**

| Reference | Proposed | Algorithm | Performance analysis parameters | Limitations |
|---|---|---|---|---|
| [13] | An efficient IDS is developed to detect network attacks in the cloud by monitoring network traffic. | Naive Bayes classifier | Average accuracy, average false positive rate and an average running time | This method was not applicable to real-time. |
| [14] | A proactive multi-cloud cooperative IDS is developed. | The deep learning algorithm, Denoising Autoencoder | Accuracy, Error rate. | This method gives better detection rate, but this has some complexity. |
| [15] | To Reduce incorrectly-classified instances an efficient IDS is developed. | Multilayer perceptron (MLP) network, and artificial bee colony (ABC) and fuzzy clustering algorithms. | MAE, RMSE, and the kappa statistic | This work fails to combine the meta-heuristic methods for system improvement. |
| [16] | An IDS for mobile clouds is developed to securely collect and fuse the data from heterogeneous client networks. | Multi-layer traffic screening and K-Means DBSCAN. | Error rate. | In feature-based traffic filtration, they fail to consider other types of traffic like HTTP and SMTP. |
| [17] | A hybrid method for an anomaly network-based IDS (A-NIDS) is presented to gain a high DR with low FPR. | Artificial Bee Colony (ABC) and AdaBoost algorithms. | Detection Rate, Accuracy, TP, TN, FP, and FN. | The efficiency of this method was improved further. |

| | | | | |
|---|---|---|---|---|
| [18] | Data Stream-based IDS to analyze the possibility of using data stream mining for enhancing the security of AMI through IDS. | Data Stream Mining algorithm. | Accuracy, Kappa Statistics, Model size, Running time, Model cost, FPR, FNR. | The accurate data stream mining algorithms to be used for the smart meter IDS in the future. |
| [19] | An IDS with the help of WKMC and the ANN classifier helps to correspond the closest cluster to the test data, according to distance or similarity measures | Weighted K-means clustering algorithm with an artificial neural network (WKMC + ANN) | Accuracy, Sensitivity, Specificity, MSE, RMSE, and MAPE. | The authentication should be enhanced more in the future. |
| [20] | An IDS or prevention system is analyzed in real-time data. | Honey pot security system | Not available | Security cost is high |
| [21] | To analyze Log Files for Postmortem Intrusion Detection. | Baum–Welch algorithm, LEarning Rules for Anomaly Detection (LERAD) | F-Measure, Recall, Precision, Time measure. | It fails to distinguish normalcy directly from the input log file where an intrusion has allegedly occurred. |
| [22] | Cross-Association for the Design of IDS helps to address the cyber and the physical dimensions. | Expectation-Maximization (EM) algorithm, cross-association algorithm. | Accuracy, the detection rate | It is not suitable for all existing IDS. |
| [23] | A Memory-Efficient Bit-Split Parallel String Matching to reduce the number of state transitions, the finite state machine tiles in a string matcher adopt bit-level input symbols. | Pattern_Partitioning, Pattern_Mapping algorithms. | Memory, the detection rate | The memory requirements should be improved more. |
| [24] | To detect the intrusion efficient system is developed using NSL-KDD CUP dataset | Joint of K-Means and KNN algorithm. | Accuracy, detection rate | It fails to detect attack in efficient manner. |
| [25] | An IDS based model helps to generate different training subsets for improving system performance. | Artificial Neural Networks and fuzzy clustering algorithm. | F-value, Recall, Precision on KDD dataset. | It fails to improve detection precision and detection stability. |

An Intrusion Detection Scheme was developed by various researchers was illustrated in the above table with its proposed technique, performance analysis, and limitations. Here, totally, thirteen works are analyzed and each paper has been used different algorithms, evaluation metrics, and advantages. However, these methods are having some drawbacks like as, fails to detect an attack inefficient manner, some methods are not applicable for real time process, and security cost is high, minimum detection accuracy and so. To overcome the problem, a new method is urgently needed. So, in this paper, an optimized FIS based intrusion detection system is proposed.

## III. OPTIMIZED FUZZY INFERENCE SYSTEM BASED INTRUSION DETECTION SYSTEM

### A. System model

The cloud-based IDS model is given in Fig 1. Nowadays cloud computing is used for a large number of the field mainly industry, governments, education, entertainment, etc. CC aims to offer suitable, on-demand, network access to a common pool of configurable computing assets, which can be quickly provisioned and discharged with negligible administration exertion or service interactions.

Nevertheless, numerous security problems emerge with the progress to this computing worldview including intrusion detection. In this paper, propose an efficient intrusion detection using OFIS. Basically, the cloud user transmits the data in the cloud or stored in the cloud. In this, every time, the received information to the cloud is captured and originality of the data will check using OFIS. This permits to identify and block intruded data while allowing access to the normal data. The proposed system consists of three modules namely; (i) PFCM based clustering, (ii) training using OFIS and (iii) testing process.

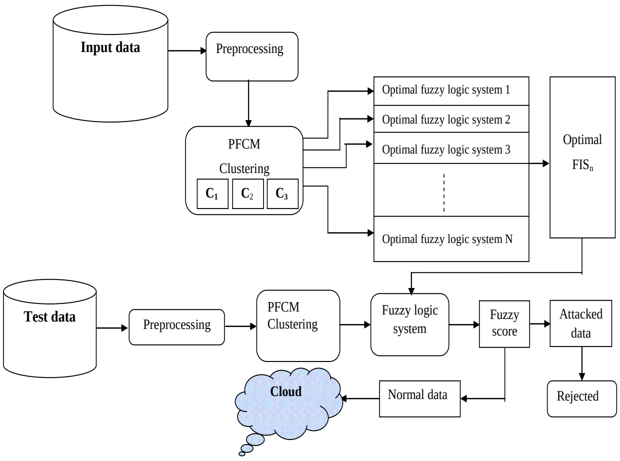


**Fig 1: Proposed intrusion detection system model**

## B. Possibilistic Fuzzy C-Means (PFCM) clustering algorithm based clustering

For intrusion detection system, in this paper NSL KDD cup dataset is utilized. The data set consist of n number of records and m number of the attribute. Each attribute has different format i.e, numeric and string. Initially, each attribute values are converted into a numeric value. This will reduce the complexity of IDS and this will make the further process easy. After preprocessing, the incoming data are given to the clustering process. Calculating intruded data in a large amount of data which creates the processing extremely complex and time-consuming. Processing, this large amount of data can give poor outcome and increase the error. To reduce this issue, in this paper the clustering method is used before ID. For clustering in this paper, the PFCM clustering algorithm is utilized. PFCM is a clustering approach which is used to group similar data based on membership function. For each cluster, PFCM creates memberships and possibilities simultaneously. PFCM is a combination of PCM and FCM that frequently eliminates different issues of FCM, PCM, and FPCM. PFCM resolves the noise sensitivity problem of FCM, concurrent clusters problem of PCM and eradicates the row sum limitations of FPCM.

Let us consider the input data $Y=(y_1,y_2,....,y_n)$. In this, we create m number of cluster $C=(c_1,c_2,....,c_m)$. The proposed PFCM optimize the following objective function which is given in equation (1).

$$O_{PFCM}(U,T,V;Y)=\sum_{k=1}^{n}\sum_{i=1}^{c}(aU_{ik}^{m}+bT_{ik}^{\eta})\times\|y_k-v_i\|_A^2+\sum_{i=1}^{c}\gamma_i\sum_{k=1}^{n}(1-T_{ik})^{\eta} \tag{1}$$

Subject to the parameters $\sum_{i=1}^{c}U_{ik}=1\ \forall k$ and $0\leq U_{ik},T_{ik}\leq 1$. Here $a>0$, $b>0$, $m>1$ and $\eta>1$. In (1) $\gamma_i>0$ is the user-specified constant. The constant $a$ and $b$ defined the fuzzy membership and typicality values in the objective function. In this equation (1), $U_{ik}$ is a membership function which is derived from the FCM. The membership function $U_{ik}$ can be calculated as follows;

$$U_{ik}=\frac{1}{\sum_{j=1}^{c}\left(\frac{\|y_k-v_i\|}{\|y_k-v_j\|}\right)^{\frac{2}{m-1}}} \tag{2}$$

Similarly, in equation (1), typically matrix $T_{ik}$ is similar to PCM. The typically matrix $T_{ik}$ can be calculated as follows;

$$T_{ik}=\frac{1}{1+\left[\frac{D^2(x_k,v_i)}{\eta_i}\right]^{1/(m-1)}} \tag{3}$$

The cluster center $V_i$ of PFCM is can be calculated as follows;

$$v_i=\frac{\sum_{k=1}^{n}(au_{ik}^{m}+bt_{ik}^{\eta})X_k}{\sum_{k=1}^{n}(au_{ik}^{m}+bt_{ik}^{\eta})},1\leq i\leq c. \tag{4}$$

The clustering process is continued on k-number of iteration. After the clustering process, the data are grouped into number of clusters. Then, the clustered data are given to the input of the optimal fuzzy Inference system to detect the data as normal or intruded data.

## C. Optimal fuzzy Inference systems for ID

In this section, IDS is attained by Mamdani fuzzy logic model. Fig 2 shows the proposed fuzzy logic model. This model makes the decision by performing the four levels namely fuzzification, generation of the fuzzy rule base, fuzzy inference system, and defuzzification. Here, each cluster assigned one FLS. The data features are given to the input of the FLS. Depend on these input variables; input crisp values of the fuzzy logic model are processed.

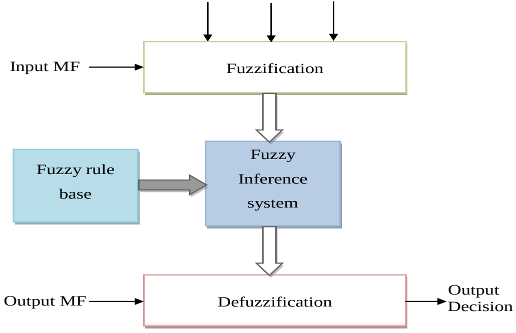


**Fig 2: The proposed fuzzy logic model**

**Fuzzification:** Input crisp values of 41 features are converted into fuzzy variables. Then, for each fuzzy variable, membership function (MF) is determined. The output parameter of this model is the trusted score. For each feature, fuzzy variables are classified in the range [0, 1] and are classified as Lowest (LL), Low (L), medium (M), high (H) and Highest (HH).

Fuzzy variables of the output probe, U2R, R2L, and DOS. For obtaining the optimum results, trapezoidal and triangular membership functions are utilized in this model. These trapezoidal and triangular membership functions are used for a boundary and intermediate variables. **Fig 3and 4** shows the membership function of fuzzy variables for the input variables and the membership function of the output variable.

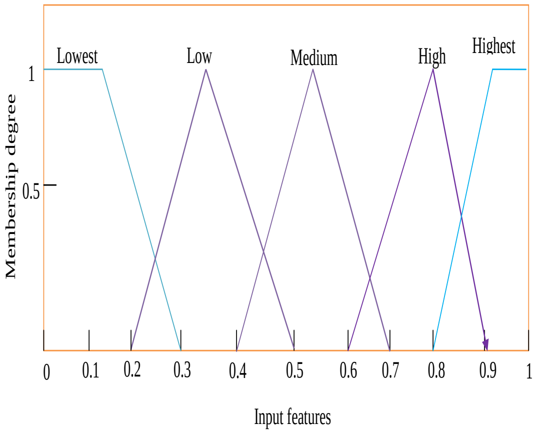


**Fig 3: Membership function of Input features**

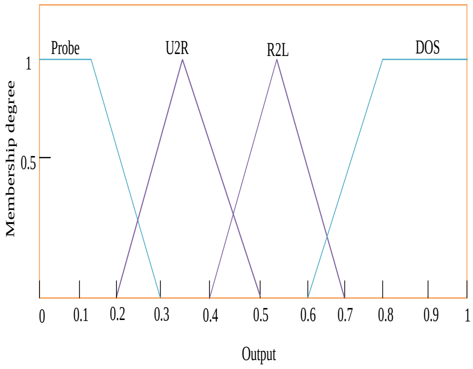


**Fig 4: Membership function of output**

**Defuzzification (Z):** For defuzzification, 5 methods are suggested to convert the fuzzy set values into crisp values [17]. The suggested defuzzification methods are Bisector of Area Method (BOA), Last of Maxima Method (LOM), Mean of Maxima Method (MOM), First of Maxima Method (FOM) and Center of gravity (COG). The sample fuzzy rules are given in table 2.

**Table 2: Sample fuzzy rule**

| Rule No. | $F_1$ | $F_2$ | ….. | $F_{41}$ | output |
|---|---|---|---|---|---|
| 1 | L | H | …. | M | normal |
| 2 | LL | M | …. | HH | probe |
| 3 | M | H | …. | LL | DOS |
| 4 | H | LL | …. | HH | Probe |
| 5 | M | H | …. | L | U2R |
| 6 | HH | L | …. | M | R2L |
| 7 | H | LL | …. | H | U2R |
| …. | …. | …. | …. | …. | …. |
| N | LL | H |  | M | DOS |

For each time, the rule base of this FIS system is to be adjusted for every time by adjusting the input and output parameters of MFs. So, it is essential to determine an optimal combination of these parameters. In this approach, the following parameters of the FIS system are to be optimized:

➢ Triangular MFs of the input variables are to be optimized. For example, if we consider a triangular shape with three peak values such as *p*, *q*, and *s* as shown in Fig5, where *q* and *s* are fixed while *p*-value is varied. In this proposed FIS system, the input variables all have triangular shapes those parameters are to be optimized. As shown in Fig 6, three parameters such as $x_1,x_2$ *and* $x_3$ are to be optimized.

➢ Among the fuzzy rules, the optimal fuzzy rules are optimally selected

➢ Along with these parameters, defuzzification methods (Z) also included finding the optimal defuzzification method.

So, the parameter of FIS selected as the initial solution. The selection of the best possible solution will provide the maximum precision rate in the network. Accordingly, so as to determine an optimal FIS and limit adjustment preliminaries time, an optimization algorithm is to be presented for FIS design issues. In this approach, GSA is presented for optimizing the parameters of the FIS system and it is described in the following section.

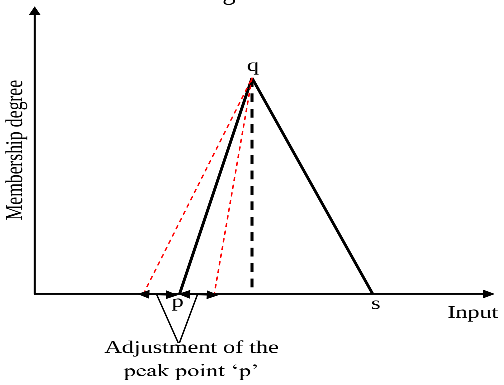


**Fig 5: Position of peak points of triangular MF**

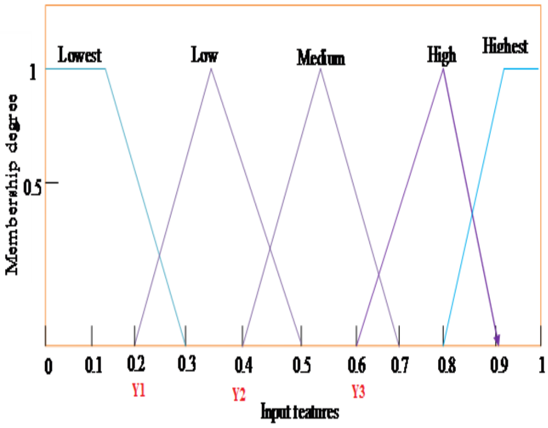


**Fig 6: Position of optimized parameters of input Feature**

In this section, elements of the FIS system are optimized using GSA for improving the process of IDS. The GSA was introduced by Rashedi in 2009 and it performs based on the behavior of Newtonian's law of gravity and motion [26].
The algorithm is planned to improve the performance in the exploration and exploitation capabilities of a populace based algorithm, based on gravity rules. The GSA algorithm based FIS system design is explained below;

**Initialization:** Initially, the FIS system is randomly initialized and parameters of GSA also initialized. Then the positions of agents or solutions are initialized with the population size N. the solution is consist of a parameter of input data (features) and n number of rules. The structure of each solution is shown in Fig 7.

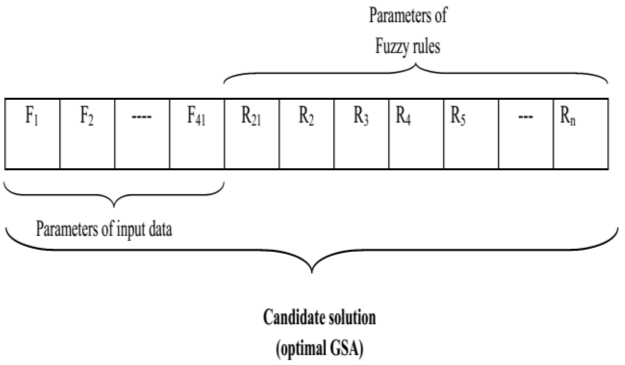


**Fig 7: The structure of the solution**

Let agents or solutions are initialized in *d* dimensional space as follows:

$$Y=\{FIS_1, FIS_2,.....,FIS_d\} \qquad (5)$$

Where, $FIS_d$ represents the position of the agent or optimal FIS system in $d^{th}$dimension. Constraints of the input parameter are given as follows:

$$\{q_m \le x_m \le s_m; \quad m=1,2,3,4 \qquad (6)$$

Input parameters $x_m$ MF is given in equations (7)

$$\begin{cases} q_m < t_m < s_m \\ q_m < q_{m+1} < s_m \\ q_{m+1} < t_{m+1} < s_{m+1} \\ s_m < q_{m+2} < s_{m+1} \\ q_{m+2} < t_{m+2} < s_{m+2} \\ s_{m+1} < q_{m+3} < s_{m+2} \\ q_{m+3} < t_{m+3} < s_{m+3} \\ s_{m+2} < s_{m+3} \end{cases} \quad for \;\; x_m \qquad (7)$$

**Fitness of each agent calculation:** After the initialization of the agent or solutions, fitness of the agent is calculated. This fitness is evaluated based on the accuracy of the system. The system with maximum accuracy value is selected as an optimal FIS system. It can be calculated as follows,

$$Fit_i = Max\{P_i\} \qquad (8)$$

Where, $P_i$denotes the precision of the $i^{th}$ solution and it can be defined as follows;

$$P=\frac{TP}{TP+FP} \qquad (9)$$

Where, TP represent the true positive and FP indicates the false positive value.

**Mass of each agent calculation:** Mass of each agent is calculated as follows;

$$M_j(k)=\frac{m_j(k)}{\sum_{i=1}^{N}(k)} \qquad (10)$$

$$m_j(k)=\frac{fit_j(k)-worst(k)}{best(k)-worst(k)} \qquad (11)$$

Where, $fit_j(k)$ represent the fitness value of the $j^{th}$ agent at time $k$. $best(k)$ Represent the best fitness value and $worst(k)$ represent the worst fitness value.

**Total force of each agent calculation:** After fitness calculation total force of each agent is calculated. It can be calculated as follows;

$$F_j(k)=\sum_{i\in kbest, i\neq j} Rand_i\, G(k)\frac{M_i(k)\times M_j(k)}{D_{j,i}(k)+\varepsilon}\left(p_i(k)-p_j(k)\right) \qquad (12)$$

Where;

$Rand_i$ →Random number in the interval [0,1]

$G(k)$→ Gravitational constant at time k

$M_i$→ Mass of $i^{th}$ agent

$M_j$ → Mass of $j^{th}$ agent $D_{j,i}$ → Distance between two agents I and j.

The gravitational constant $G(k)$can be calculated as follows;

$$G(k)=G_0\times\exp\left(-\beta\times\frac{k}{k_{\max}}\right) \qquad (13)$$

Where,

$G_0$ → Initial value

$\beta$ → Constant value

$k$ → Current iteration

$k_{\max}$ → Maximum number of iteration

**Acceleration of each agent calculation:** Acceleration of each agent can be calculated as follows;

$$A_j(k)=\frac{F_j(k)}{M_j(k)} \qquad (14)$$

**Updation using GSA:** The position updation is an important process for the optimization algorithm. To calculate the efficient agent, updation function is utilized. After calculating the fitness to the position of the agents, it will be updated to the next position.

The agent velocity and position Updation function is given in equation (15) and (16).

$$V_i\ (k+1) = rand_i \times V_i\ (k) + a_i\ (k) \tag{15}$$

$$Y_i\ (k+1) = Y_i\ (k) + V_i\ (k+1) \tag{16}$$

Where,

$rand_i$ represent the arbitrary number between interval [0,1].

$V_i\ (k+1)$ is the velocity of $i^{th}$ the agent during $k^{th}$ the iteration and $Y_i\ (k+1)$ is the position of $i^{th}$ an agent at $k^{th}$ iteration.

**Termination:** Above operations are continued until finding the optimal solution or optimal FIS system. Once the optimized FIS system is obtained, then the algorithm will be terminated. This optimized FIS system is used as the intrusion detection system which increases the total accuracy of the system.

### D. Testing process

After the training process, the test case data are tested in which the data are classified as probe attack, U2R attack, R2L attack, DOS attack, and normal data. Here, initially, the test data are given as the input. After that, the data corresponding cluster-based FIS system is selected. Then, the trained FIS system is assigned to input data. Based on the input data FIS system finally produced the output score value ($O^{Score}$). Based on the score value the data is classified as five classes namely, probe attack, U2R attack, R2L attack, DOS attack and normal data. For classification, in this paper, four threshold values are assigned namely, $Th_1$, $Th_2$, $Th_3$ and $Th_4$. These threshold values are fixed based on the data score value ($O^{Score}$ ). The classification is performed using equation 17.

$$output = \begin{cases} 0 < O^{score} < Th_1;\ \ DOS\ attack\ adta \\ Th_1 \le O^{score} < Th_2;\ probe\ attack\ data \\ Th_2 \le O^{score} < Th_3\ ; U2R\ attack\ data \\ Th3 \le O^{score} < Th_4\ ; R2L\ attack\ data \\ Th4 \le O^{score}\ ;\ Normal\ data \end{cases} \tag{17}$$

**Table 3: Optimization of the FIS system using GSA**

```
Algorithm 1: Optimization of the FIS system using GSA
Input: Clustering data, parameter of GSA , Parameter of FIS system
Output: Optimized FIS

1. Initialize the position of agents or solutions (FIS system).
2. While (k< Max Generation) or (stop criterion)
3. Calculate fitness for using equation (5).
4. For  i=1 to N do
        Update G(k), best (k), worst(k) and $M_i(k)$
5. End for
6. Calculate total force of each agent using equation (12)
7. Calculate the acceleration  of each agent using equation (14)
8. Velocity of each agent is updated using equation (15)
9. Position of each agent is updated using equation (16)
10.   End while
11.   Steps 3-9 are continued until finding the optimal solution or Optimized FIS system
12.   End
13.   This optimized FIS system is used as the intrusion detection process
```

## IV. RESULT AND DISCUSSION

The proposed IDS experimental results are analyzed in this section. The proposed methodology implemented in JAVA with Cloud Sim tools and a series of experiments were performed on a PC with Windows 7 Operating system at 2 GHz dual-core PC machine with 4 GB main memory running a 64-bit version of Windows 2007. To evaluate the performance of the proposed PFCM+OFIS based intrusion detection method, a series of experiments on the NSL KDD CUP1999 dataset were conducted.

### A. Dataset Description

The NSL-KDD data set is a refined version of its predecessor KDD"99 data set and this dataset are widely applied for the intrusion detection system. This dataset contains five million records and each record consists of 41 features. The attack classes present in the NSL-KDD dataset are grouped into four classes namely Probe attacks, U2R attacks, R2L attacks and DoS attack. In this dataset, the class attribute is consisting of the binary value. The dataset consists of the only binary class attribute. Here, training and testing datasets are available.

### B. Evaluation metrics

The evaluation of the suggested intrusion detection system is carried out applying the following metrics as proposed by equations, given below:

**Precision:** It is measured ratio of the number of normal data inquired to the total number of normal and abnormal data detected which is afforded in equation (18).

$$P = \frac{TP}{TP+FP} \tag{18}$$

**Recall:** It is measured ratio of the number of normal data inquired to the total number of data present in the dataset which is afforded in equation (19).

$$R = \frac{TP}{TP+FN} \tag{19}$$

**F-measure:** F-measure is determined as the harmonic mean of precision and recalls metrics which is afforded in equation (20).

$$F = \frac{2PR}{P+R} \qquad (20)$$

### C. Simulation results

The main aim of this study is to detect an intruded data from incoming data using OFIS which is used to increase the security of the system.

The simulation is done working platform of JAVA with Cloud Sim tools. The proposed methodology test bed is given in Fig 8. Moreover, Fig 9-11 shows the simulation results. The main aim of this study is to detect an intruded data from incoming data using OFIS which is used to increase the security of the system. The simulation is done working platform of JAVA with Cloud Sim tools. The proposed methodology test bed is given in Fig 8. Moreover, Fig 9-10 shows the simulation results obtained from the proposed intrusion detection system.

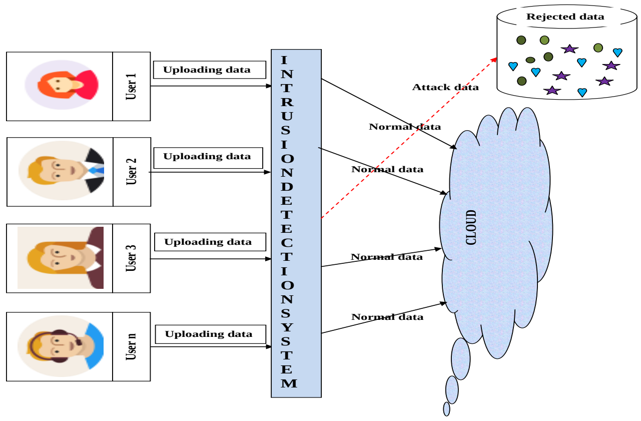
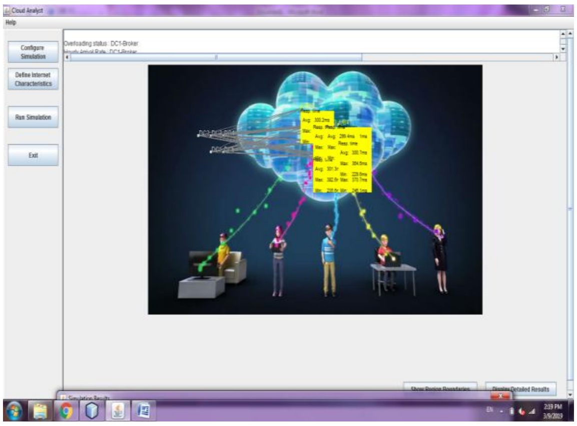


Fig 8: Testbed of proposed approach

Fig 9: Simulation started window



Overloading status : DC1-Broker
Hourly Arrival Rate : DC1-Broker

Fig 10: Simulated window

### D. Performance analysis based on different classifier

In this section, the experimental results attained from the proposed intrusion detection system are explained. In this suggested approach, at first, the data are preprocessed. Then the preprocessed data are afforded to the clustering process. For the clustering process, in this research PFCM is utilized. Then the clustered data are given to the classification process. For classification OFIS system is utilized. In this stage, the data are classified as four types namely, normal, probe, DOS, U2R and R2L. To prove the effectiveness of the proposed methodology, the proposed algorithm compared with different classifier namely, naïve bayes, artificial ANN, K-nearest neighbour (KNN) and FLS. The performance is analyzed in terms of precision, recall and F-measures. Fig 12-14 gives the performance result of proposed approch.

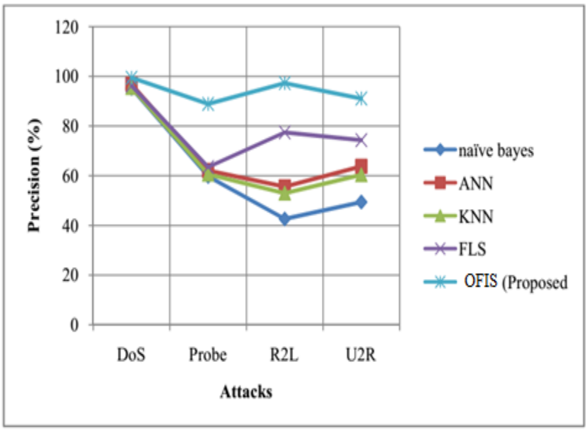


Fig 12: Comparison of precision of different classification techniques for various attacks

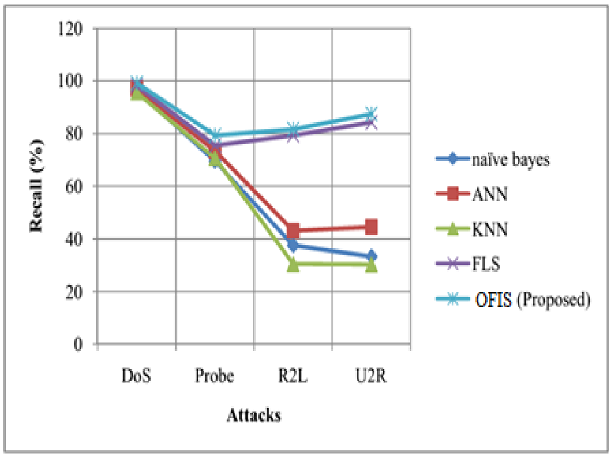


Fig 13: Comparison of recall of different classification techniques for various attacks

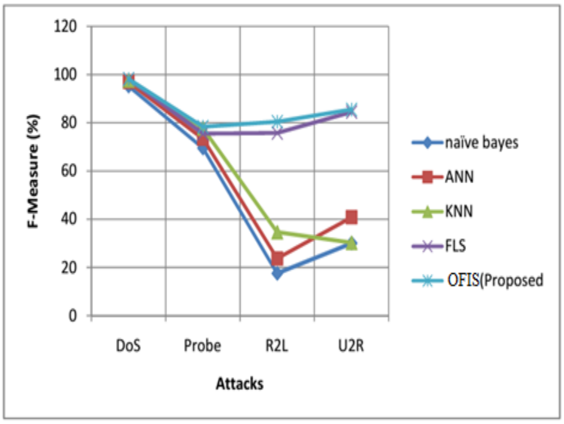


**Fig 14: Comparison of F-measure of different classification techniques for various attacks**

Fig 12 shows the comparative result of proposed approach based on precision measure for varies attacks. Here, we compare our proposed technique performance with different algorithms. When analyzing Fig 12, our proposed method attains the maximum precision of 99.45 % for Dos attack detection, 88.93% for probe attack detection, 97.28% for R2L attack detection and 91.12% for U2R attack detection. Although, proposed OFIS based intrusion detection system attain better precision compare to other classifier. Due to the optimization of fuzzy parameters, decision making the performance of Fuzzy system is improved further so that the classification precision is increased. Moreover, in Fig 13, comparison of recall measure of different classification technique for vary attack is shown.

In Fig 13, the x-axis shows the different types of attack and the y-axis shows the recall value. When analyzing Fig 13, based IDS and OFIS based IDS provide a better recall value compare to naïve Bayes based IDS, ANN-based IDS, and KNN based IDS. Compare to FLS based IDS, OFIS based approach is attained the better result. This is because of, the optimal FIS. Similarly, in Fig 14, the performance of the proposed approach is analyzed in terms of F-measure. Here, also our proposed approach attains better results. From the result section, we clearly understand, our proposed method attains a better result compared to the different classifier.

### E. Performance analysis based on clustering algorithm

In this paper, to reduce the complexity, time-consuming and increase the detection accuracy, before ID, the data are clustered using the PFCM clustering method. PFCM is a hybridization of FCM and PCM algorithm. The performance-based on clustering algorithm is analyzed in this section.

**Table 2: Performance analysis based on the precision measure by varying cluster size**

| Cluster size | PFCM | FCM | PCM | K-means clustering |
|---|---|---|---|---|
| 3 | 95.35 | 90.56 | 91.45 | 85.28 |
| 4 | 98.45 | 91.47 | 92.69 | 86.58 |
| 5 | 97.56 | 89.57 | 90.57 | 83.19 |
| 6 | 96.75 | 87.28 | 89.67 | 81.79 |

**Table 3: Performance analysis based on Recall measure by varying cluster size**

| Cluster Size | PFCM | FCM | PCM | K-means clustering |
|---|---|---|---|---|
| 3 | 90.62 | 85.78 | 87.18 | 80.35 |
| 4 | 94.72 | 86.53 | 88.34 | 81.37 |
| 5 | 92.28 | 84.18 | 86.29 | 79.47 |
| 6 | 93.67 | 83.78 | 85.83 | 78.38 |

**Table 4: Performance analysis based on F-measure measure by varying cluster size**

| Cluster Size | PFCM | FCM | PCM | K-means clustering |
|---|---|---|---|---|
| 3 | 91.56 | 86.56 | 88.79 | 80.21 |
| 4 | 95.57 | 87.29 | 89.92 | 81.52 |
| 5 | 93.78 | 85.21 | 88.69 | 79.32 |
| 6 | 92.68 | 84.32 | 87.35 | 78.32 |

The above table 2-4 shows the performance of the proposed methodology based on the clustering algorithm. Here, by varying cluster size, we measure the performance in terms of precision, recall, and F-measure. In table 2, our proposed PFCM based clustering algorithm attains a better result compared to other clustering algorithms. Here, we attain the maximum precision of 98.45% which is high compared to other algorithms. Similarly, in table 3 and 4 also, we obtain a better result. This is because of PFCM clustering. The PFCM overcome the difficulties present in the FCM and PCM. So, in this paper PFCM based method attain better results. From the result, we clearly understand our proposed method attain a better result compare to another method because of clustering and optimal FIS system.

### F. Comparison with published papers

To prove the effectiveness of proposed methodology, in this paper, we compare our performance of proposed methodology with existing works namely, PCA+NN [29], MLP [28], ABC+FCM+NN [15] and PFCM+RNN [27]. In [29], a combination of principle component analysis and neural network based intrusion detection is made. For feature selection they utilized PCA and classification they utilized the artificial neural network. In [28], Multi Layer Perceptron (MLP) is used for IDS which is based on off-line analysis approach. The hybridization of a Multilayer Perceptron (MLP) network, artificial bee colony (ABC) and fuzzy clustering algorithms based IDS is developed in [15]. Similarly, in [27], the IDS are developed based on PFCM with RNN. For comparing these methods, NSL-KDD cup 99 dataset is utilized. Comparative analysis based on the precision measure for KDD CUP 99 dataset is given in Fig 12.

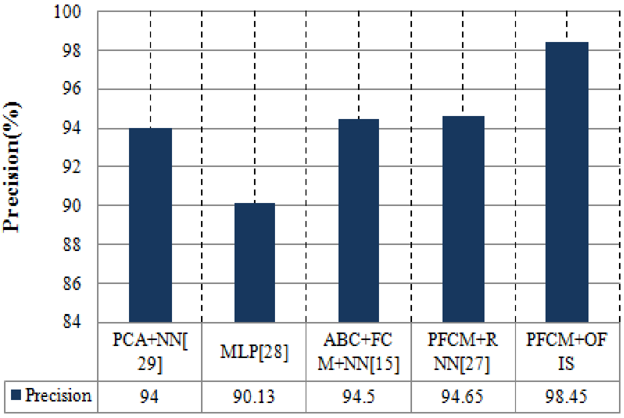


**Fig 12: Comparative analysis based on the precision measure for KDD CUP 99 dataset**

OFIS+PFCM based IDS is explained in this paper. Here, for clustering process, PFCM is utilized for clustering and OFIS is used for ID. When analyzing Fig 12, we obtain the average maximum accuracy of 96.54% which is 94% for using PCA-NN [29], 90.13% for using MLP based IDS[28], 94.5% for using [15] and 94.65% for using PFCM+RNN [27] and 98.45% for PFCM+OFIS.From the result, we clearly understand our proposed approach is better when compared to other approaches.

## VI. CONCLUSION

Nowadays, system security is one of the major worries because of different attacks and vulnerabilities in the cloud. As a result, attack detection is an imperative segment in system security. A hybridization of PFCM and OFIS generates new IDS which are presented in this paper. The difference between normal and abnormal data is done by our proposed intrusion detection system. The OFIS is generated using GSA. The experimental results using the NSL KDD CUP 1999 dataset demonstrates the effectiveness of our approach which provides better precision than the existing method. In future, we will improve the security of the data using cryptographic algorithms.

## REFERENCES

1. Anthony T. Velte, Toby J. Velte and Robert Elsenpeter, “Cloud Computing – A Practical Approach”, Tata McGrawHill Edition, ISBN: 978-0-07-162695-8.
2. Mell, Peter, and Tim Grance, “Effectively and securely using the cloud computing paradigm” NIST, Information Technology Lab 2009.
3. Adel NadjaranToosi and Mohsen Kahani, "A new approach to intrusion detection based on an evolutionary soft computing model using neuro-fuzzy classifiers",journal of computer communications, vol. 30,pp. 2201–2212, 2007.
4. H. Debar, M. Dacier and A. Wespi, “Towards a taxonomy of intrusion detection systems”, journal of computer networks, vol. 31, pp.805–822, 1999.
5. Chavan, S.; Shah, K.; Dave, N.; Mukherjee, S.; Abraham, A.; Sanyal, S.; , "Adaptive neuro-fuzzy intrusion detection systems," Proceedings. ITCC 2004. International Conference on Information Technology: Coding and Computing, 2004., vol.1, no., pp. 70- 74 Vol.1, 5-7 April 2004.
6. P. Kachurka, V.Golovko, “Neural Network Approach to Real Time Network Intrusion Detection and Recognition". Proceedings of the 6th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications, Prague, 15-17 September 2011.
7. Fengxi Song; ZhongweiGuo; Dayong Mei; , "Feature Selection Using Principal Component Analysis,", 2010 International Conference on System Science, Engineering Design and Manufacturing Informatization (ICSEM), vol.1, no., pp.27-30, 12-14 Nov. 2010.
8. Dae-Ki Kang, Fuller and Honavar, “Learning classifiers for misuse and anomaly detection using a bag of system calls representation”, Proceedings from the Sixth Annual IEEE SMC Information Assurance Workshop, pp. 118- 125, 2005.
9. Chirag Modi, Dhiren Pate, "A Survey of Intrusion Detection Techniques in Cloud", Elsevier Journal Of Network And Computer Applications, Vol 36, No 1, Pp. 42–57, Jan 2013.
10. Kaihwang, "Trusted Cloud Computing with Secure Resources and Data Coloring", In Proceedings of IEEE Transaction of Internet Computing, Vol. 14, No. 5.
11. Swati Ramteke, Rajesh Dongare and KomalRamteke, "Intrusion Detection System for Cloud Network Using FC-ANN Algorithm", International Journal of Advanced Research in Computer and Communication Engineering vol. 2, no.4, 2013
12. L. Huang, “A new approach to intrusion detection using Artificial Neural Networks and fuzzy clustering”, School of Management, Fudan University, Shanghai 200433, PR China, Department of Information Systems, City University of Hong Kong, Tat Chee Avenue, Kowloon, Hong Kong.
13. Idhammad, Mohamed, Karim Afdel, and Mustapha Belouch, "Distributed intrusion detection system for cloud environments based on data mining techniques," Procedia Computer Science, Vol. 127, No. 35-41, 2018.
14. Abusitta,Adel, Martine Bellaiche, Michel Dagenais, and TalalHalabi, "A deep learning approach for proactive multi-cloud cooperative intrusion detection system," Future Generation Computer Systems, 2019.
15. Bahram Hajimirzaei and NimaJafariNavimipour, “Intrusion detection for cloud computing using neural networks and artificial bee colony optimization algorithm”, Journal of ICT Express, 2018
16. Hajimirzaei, Bahram, and NimaJafariNavimipour, "Intrusion detection for cloud computing using neural networks and artificial bee colony optimization algorithm," ICT Express, Vol. 5, No. 1, pp. 56-59, 2019.
17. MehrnazMazini, BabakShirazi and IrajMahdavi, “Anomaly network-based intrusion detection system using a reliable hybrid artificial bee colony and AdaBoost algorithms”, Journal of King Saud University - Computer and Information Sciences, 2018
18. Mustafa Amir Faisal, Abel Sanchez, “Data-Stream-Based Intrusion Detection System for Advanced Metering Infrastructure in Smart Grid: A Feasibility Study,” IEEE Systems Journal, 2014.
19. RafathSamrin and DevaraVasumathi, “Hybrid Weighted K-Means Clustering and Artificial Neural Network for an Anomaly Based Network Intrusion Detection System”, Journal of intelligence system, 2016
20. MuhammetBaykara and Resul Das, "A novel honeypot based security approach for real-time intrusion detection and prevention systems", journal of information security and application, vol.41, pp.103-116, 2018
21. Karen A. Garc Monroy, Luis A. Trejo, “Analyzing Log Files for Postmortem Intrusion Detection,” IEEE Transactions on Systems, man and Cybernetics, Volume. 42, No. 6, November 2012.
22. Piroska Haller, Bela Genge, “Using Sensitivity Analysis and Cross-Association for the Design of Intrusion Detection Systems in Industrial Cyber-Physical Systems,” IEEE Translations and content mining, 2016.
23. Hyun Jin Kim, Hong-Sik Kim, “A Memory-Efficient Bit-Split Parallel String-matching Using Pattern Dividing for Intrusion Detection Systems,” IEEE transactions on parallel and distributed systems, vol. 22, no. 11, 2011
24. AboosalehM.Sharifi, Saeed K. Amirgholipour and AlirezaPourebrahimi, “Intrusion Detection Based on Joint of K-Means and KNN”, Journal of Convergence Information Technology, vol.10, no.5, 2015
25. Gang Wang, Jinxing Hao, Jian Ma and Lihua Huang, “A new approach to intrusion detection using Artificial Neural Networks and fuzzy clustering”, journal of Expert Systems with Applications, vol. 37, pp.6225–6232, 2010.
26. EsmatRashedi, Hossein Nezamabadi-pour and SaeidSaryazdi, "GSA: A Gravitational Search Algorithm", Information Sciences, vol. 179, no. 13, pp. 2232-2248, 2009

27. M. Manickam, N. Ramaraj and C. Chellappan, "A Combined PFCM and Recurrent Neural Network based IDSfor Cloud Environment", International Journal of Business Intelligence and Data Mining, vol.1, no.1, 2017
28. Mehdi moradi and Mohammad zulkernine, "A Neural Network Based System for Intrusion Detection and Classification of Attacks ", IEEE international conference on advances in intelligent system, 2004
29. Zeenat Mahmood, Chetan Agrawal, Syed Shadab Hasan and Syeda Zenab, "Intrusion Detection in Cloud Computing Environment using Neural Network", International Journal of Research in Computer Engineering and Electronics, vol.1, no.1, 2014.

## AUTHORS PROFILE

**S. Immaculate Shyla** received her MCA degree from Anna University of Technology, Tirunelveli, in 2011. She has received her M.Phil degree in Computer Science from Manonmaniam Sundaranar University, Tirunelveli, in 2013.She is working as an Assistant Professor in Department of Computer Science at St.Alphonsa College of Arts and Science, Nagercoil, India. Currently, she is a PhD candidate at Manonmaniam Sundaranar University, Tirunelveli Her research interests include Cloud Security, Intrusion Detection in Cloud and Fuzzy Logic.

**Dr.S.S.Sujatha**, received her MCA degree from Alagappa University, Karaikudi, in 1993. She has received her M.Phil degree in Computer Science from Manonmaniam Sundaranar University, Tirunelveli, in 2003 and Ph.D in Computer Science from Mother Teresa Women's University, Kodaikanal. She is working as an Associate Professor in the Department of Computer Applications at S.T.Hindu College, Nagercoil since from 1994. Her research interests include Digital Image Processing, Digital Watermarking and Cloud Security. She has published 20 papers in reputed International Conferences and International Journals.