Journal of Theoretical and Applied Information Technology
10th March 2014. Vol. 61 No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

A FUZZY NEURAL NETWORK AND MULTIPLE KERNEL FUZZY C-MEANS ALGORITHM FOR SECURED INTRUSION DETECTION SYSTEM

P.ANANTHI[1], P.BALASUBRAMANIE[2]
[1] Assistant Professor, Kongu Engineering College, India;
[2]Professor, Kongu Engineering College, India.
[1]Email: ananthi.scholar@gmail.com

ABSTRACT

An Intrusion Detection System (IDS) is a security layer used to detect constant intrusive behavior in information systems. Many intrusion detection systems have been proposed based on the various data mining approaches such as decision tree, clustering, etc. Although the intrusion detection system is efficient way to find the attacks in the system, existing ones have some disadvantages which affects the performance of the system. It is observed that Neural Networks improves the overall performance of the intrusion detection system when it is integrated with a clustering approach. This research work aims to improve the performance of the intrusion detection system through the application of Fuzzy Neural Network along with an efficient fuzzy clustering method. In this proposed approach, initially Multiple Kernel Fuzzy C-Means (MKFCM) technique is used to construct different training subsets. The performance of the fuzzy clustering approach is improved through MKFCM. Then, different FNN models are trained to formulate different base models according to the different training models. Then the final results are aggregated through fuzzy based approach. The performance of the proposed MKFCM-FNN approach is compared with other existing approaches.

Keywords: *Intrusion Detection System, Fuzzy Neural Network, Multiple Kernel Fuzzy C-Means, false positive*

1. INTRODUCTION

In networks, many new technologies and heuristic communication equipments such as cheap, low-power, multifunctional devices has been introduced which is widely used. A network is defined as a group of tiny nodes which is placed in unattended environment. The network structure can be either structured or non-structured pattern and it does not contain any predefined structure.

For the past years, the network security plays an vital role in malicious traffic detection because the traffic are generated by network attacks, such as denial of service (DoS), viruses, worms, trojan horses, spyware, and so on. Furthermore, malicious traffic makes network performance ineffective and dilemma users [1].

Nowadays the attacks in the network turn out to be unavoidable, the security systems such as firewalls cannot identify the powerful attacks such as denial of service, viruses, worms etc so that the security systems like intrusion detection system, prevention system are used over there. These prevention systems are better than the common security system because it detect the powerful attacks, computer system usage etc [2, 3]. In order to increase the strength of the system security the intrusion detection system has common security tools. Many intrusion detection systems is introduced based on the statistical algorithm, heuristic algorithm and many researches has been conducting for improving the security solutions [4-7]. The CIA guarantee has been violated or intruded by a group of attackers and attempt to exploit the security, integrity, confidentiality, and availability. It is usually called as an intrusion. Even though it protects the powerful attacks from the group of users in network, it has some problems in protecting system.

Intrusion detection system has some of the predefined actions for attacks or intrusions in the network [8]. IDS functions includes analyzing the system and networks activities, typical pattern of the each attacks are identified, identifying the user policy violations, protecting the system and file integrities are some of the main functions of this system.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

The detection system is build based on the data mining techniques and it is used mainly to improve the security of the system. It also has some disadvantages such as special requirements of the IDS are not fulfilled. Some of the researchers introduced game theory for improving and building the IDS system. In these systems the game strategies has been used to detect the attackers [9].

Many research works and studies have been conducted based on the intrusion detection system to improve the security of the system. In recent years, ANN based IDS is observed to provide good results and security. But, the ANN based IDS has certain drawbacks such as lower detection precision, especially for low-frequent attacks and detection in stability [10].

ANN integrated with fuzzy clustering has been presented to overcome the drawbacks of the ANN based IDS [11]. The present research work develops an extension of the FC-ANN approach. In order to overcome the drawbacks of fuzzy clustering, an efficient Multiple kernel Fuzzy C-Means clustering approach is presented in this research work. Moreover, in this research work, Fuzzy Neural Network (FNN) is used for better performance.

2. LITERATURE SURVEY

A number of research works have been carried out for improving the WSN and its application. Most of the network applications are unprotected and open as a result it can be attacked by group of people. IDS has been developed for detecting the attacks in networks and used to protect the system or networks from the attackers.

Wu and Banzhaf [2] studied and evaluated the IDS performance based on the above mentioned algorithm and its problems. In order to reduce FP's based on the fuzzy alert correlation by using the combination of IDS and fuzzy inference system [12] and [13]. Although the above mentioned system reduce the FP's it is not reducing FN's. To overcome this problem Sourour et al. in [14] studied about a system based on the environmental awareness of intrusion detection and prevention system and it is also used to detect the FN's.

Attack Session Extraction (ASE) was presented in [15] to build a group of traffic traces gives potential FPs and FNs to IDSs. Later the ASE was developed for a bigger system, called the PCAPLib system is studied in [16]. The PCAPLib system not only hauled out and categorizes the real-world traffic confine from Campus BetaSite [17] into proper categories by leveraging multiple IDSs. On the other hand, prior work only paying attention on investigation how to reduce FPs and/or FNs in IDSs.

Based on the intrusion detection system and it functions that IDS will affect their detections and lifetime of the networks. In [18] Lightweight Intrusion Detection is used to overcome the problems of the traditional intrusion detection system. In [19] a new intrusion detection system based on the ontology is presented to protect the network from various attacks.

The processing of the data in the critical infrastructure is mainly uses the wireless sensor network. Coppolino et al [20] presented an intrusion detection system and it is mainly used for detecting the malicious activities in Critical Infrastructure (CI) and it parts. The improved and secured communication in CI's, proper remediation/reconfiguration actions, are some of the results during the usage of the IDS. The simulation results of this approach are evaluated against the serious attacks such as sinkhole, bogus packet. It also satisfies the current state of the requirement.

In recent times due to the development of the wireless technologies usage of the wireless sensor technologies also increased from academic to military. These development leads to the malicious activities and attacking the networks because most of the networks are used in open environments. Traditional security systems are mainly used for the secure communication. Because of these developments in communication traditional security systems are cannot be developed. Gurdip Kaur et al [21] presented a intrusion detection system based on the honeypot and swarm intelligence is proposed. Sometimes the denial service attack causes the false alarm detection rate which can be solved by using this framework and detecting the intruder. The pattern recognition method is used for finding the future attacks and proposed framework is evaluated based on the speed and accuracy of the intruder detection.

3. METHODOLOGY

This research work presents an improved version of the intrusion detection system based on the Fuzzy C- Means clustering along with the Fuzzy Neural Network. The proposed system mainly used for detecting the malicious activities and it consist of the multiple Kernals Fuzzy C-Means clustering, FNN module and fuzzy aggregation module.

Journal of Theoretical and Applied Information Technology
10th March 2014. Vol. 61 No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

3.1 Framework Of Ids Based On Ann And Fuzzy Clustering

The fuzzy clustering is used to cluster the given dataset and the given dataset are trained by using the fuzzy neural network. The final results are obtained by using the membership grades along with the new artificial neural network [22]. The fig 1 describes the framework of the proposed system.

The proposed framework includes both training phase and testing phase; the training phase is explained in the following steps

Step 1: The arbitrary dataset DS is divided into training set TR and testing TS. Using the Kernals Fuzzy C-Means clustering model the dataset is divided into different training datasetsTR1; TR2; …; TRk.

Step 2: For each training subset TR_i (i = 1; 2; …; k), the FNN model, FNN_i, (i = 1; 2; …; k) is training by the specific learning algorithm to formulate k different base FNN models.

Step 3: In order to reduce the error for every FNN_i, simulate the FNN_i using the whole training set TR and get the results. Then we use the membership grades, which were generated by fuzzy clustering module, to combine the results. Subsequently, train another new FNN using the combined results.

The input is directly given to the fuzzy neural network along with the multiple Kernals Fuzzy C-Means module and final results are obtained by using fuzzy aggregation module.

3.2 MULTIPLE KERNELS FUZZY C-MEANS CLUSTERING

3.2.1 Objective Function

In kernel methods, the map features are mainly used for the mapping the features of the data to new feature space and used to discover the new relationship between the data [23]. Consider mappings as a $\Phi = \{\varphi 1, \varphi 2, \cdots, \varphi M\}$. The k^{th} mapping is defined as φk which is mainly used for recodes the l-d data **x** as a vector φk (**x**) and its dimensionality is mentioned asLk. The mercer kernels is defined as $\{\kappa 1, \kappa 2, \ldots, \kappa M\}$ which is related to implicit mappings.

$$k_k(X_i, X_j) = \phi_k(X_i)^T \phi_k(X_j)$$

The final results obtained from the above equation is still satisfies the mercer condition and the non negative combination of these features maps is defined as φ′

$$\phi'(x) = \sum_{k=1}^{M} \omega_k \, \phi_k(x), with \; \omega_k \geq 0$$

Although the result obtained is an efficient one the implicit mapping of the above equation does not have same dimensionality such as linear combination is possible. The new independent mappings are defined as $\Psi = \{\psi 1, \psi 2, \ldots, \psi M\}$ from the original mappingΦ.

$$\psi_1(x) = \begin{bmatrix} \phi_1(x) \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \psi_2(x) = \begin{bmatrix} 0 \\ \phi_2(x) \\ \vdots \\ 0 \end{bmatrix}, \ldots, \psi_M(x) = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ \phi_M(x) \end{bmatrix}$$

The new defined independent mappings are used to convert x into L-d vector where $L = \sum_{k=1}^{M} L_k$. The orthogonal bases are created by connecting the dimensions between the features because the feature spaces have infinite dimensionalities. The orthogonal bases created are used to construct the new mappings by ensuring the same dimensionality. The new set of orthogonal bases are defined as a

$$\psi_k(X_i)^T \psi_k(X_j) = k_k(X_i, X_j)$$

$$\psi_k(X_i)^T \psi_{k'}(X_j) = 0, if \; k \neq k'$$

The orthogonal base created is used to avoid the cross terms between the different implicit mappings and the inner product of the data is evaluated by the kernel methods. The non negative linear expansion is defined as $\psi(\boldsymbol{x}) = \sum_{k=1}^{M} \omega_k \, \psi k \, (\boldsymbol{x})$, and used to mapping the data into an implicit feature space. The objective function defined as

$$J(w, U, V) = \sum_{i=1}^{N} \sum_{c=1}^{C} u_{ic}^{m} (\psi(X_i) - v_c)^T (\psi(X_i) - v_c) \quad (1)$$

$$\psi(x) = \omega_1 \psi_1(x) + \omega_2 \psi_2(x) + \cdots\cdots + \omega_M \psi_M(x) \; subject \; to \; \omega_1 + \omega_2 + \cdots + \omega_M = 1$$

$$and \; \omega_k \geq 0 \; \forall k$$

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

$$and \sum_{c=1}^{C} u_{ic} = 1 \ \forall i$$

$$and \ u_{ic} \geq 0 \ \forall i, c$$

$$and \sum_{i=1}^{N} u_{ic} > 0 \ \forall c$$

Where

v_c c^{th} cluster center in the implicit feature space.

$\boldsymbol{w} = (\omega_1, \omega_2, \ldots, \omega_M)^T$ Vector consisting of weights.

$\boldsymbol{U}$ an N × C membership matrix whose elements are the memberships u_{ic}

$\boldsymbol{V}$ an L × C matrix whose columns correspond to cluster centers.

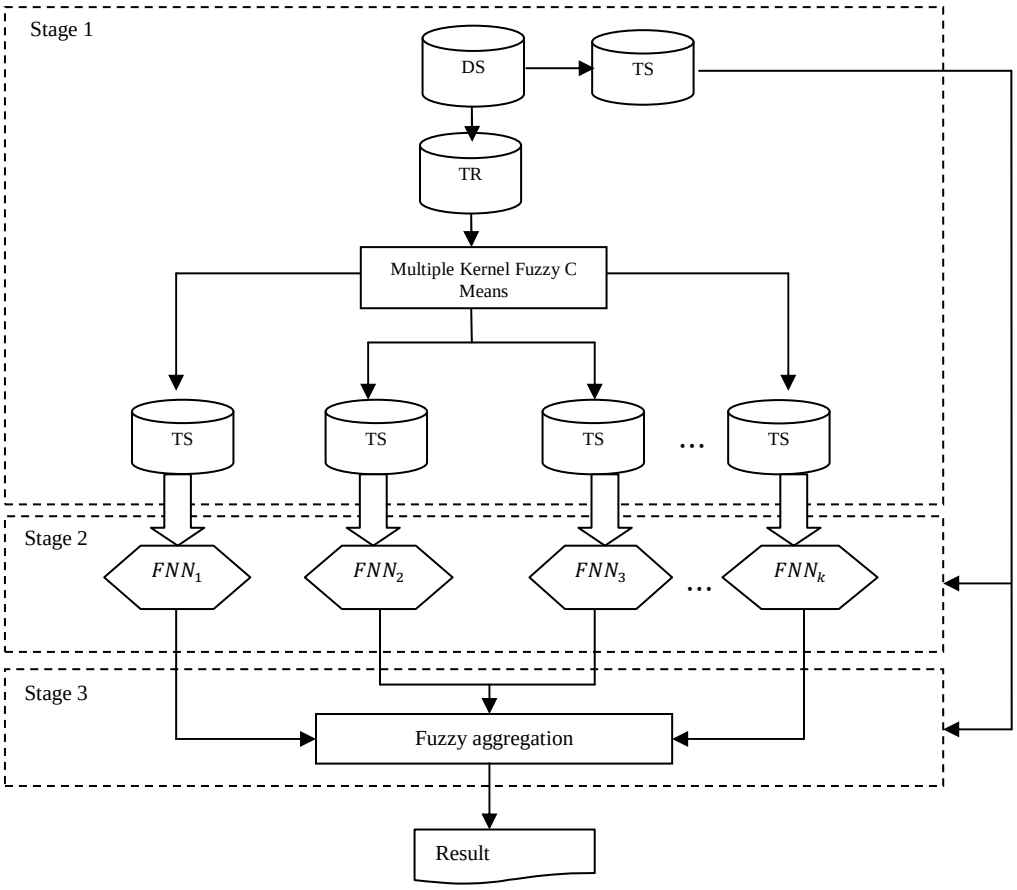

Figure 1: Framework of MKFCM-FNN for IDS

3.2.2 Optimizing Memberships

The main aim of the Multiple Kernals Fuzzy C-Means clustering is used to minimize the objective function (1) by finding the values of the parameters such as weights$\boldsymbol{w}$, memberships $\boldsymbol{U}$ and cluster center$\boldsymbol{V}$. The cluster centers in the implicit feature cannot be evaluated directly because the implicit feature space is not accessible for MKFC. The distance between data and cluster center can be calculated by using the formula $D_{ic}^2 = (\psi(\boldsymbol{x_i}) - \boldsymbol{v}c)T\,(\psi(\boldsymbol{x_i}) - \boldsymbol{v}c)$ where D_{ic} is the distance between data and cluster center, $\boldsymbol{x_i}$ denotes the data in the cluster and cluster center isv_c. Then the equation (1) is defined as

$$J(w, U, V) = \sum_{i=1}^{N} \sum_{c=1}^{C} u_{ic}^m D_{ic}^2 \qquad (2)$$

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

The function can be calculated by using the parameters such as weight and cluster have same value, distance is constant and constraint is defined as $t\sum_{c=1}^{C} u_{ic} = 1$ then the function is defined as a

$$J_\lambda(U,V) = \sum_{i=1}^{N}\sum_{c=1}^{C} u_{ic}^m D_{ic}^2 + \lambda\left(\sum_{c=1}^{C} u_{ic} - 1\right)$$

The functions derivatives are defined as zero with respect to the membership functions u_{ic} and for each membership u_{ic} define it as a

$$\frac{\partial J_\lambda}{\partial u_{ic}} = mD_{ic}^2 u_{ic}^{m-1} + \lambda = 0$$

The solution for u_{ic}, is

$$u_{ic} = \left(\frac{-\lambda}{m}\right)^{\frac{1}{m-1}} \frac{1}{D_{ic}^{2/(m-1)}}$$

The efficient and closed form of results are obtained for the problem of optimal membership function in multiple kernel fuzzy c means clustering algorithm with the help of constraint $\sum_{c=1}^{C} u_{ic} = 1$, and can eliminate the value of λ

$$u_{ic} = \frac{1}{\sum_{c'}^{C}\left(\frac{D_{ic}^2}{D_{ic'}^2}\right)^{\frac{1}{m-1}}} \quad (3)$$

3.2.3 Optimizing Weights

From equation (3) the optimal membership $\boldsymbol{U}$ can be finding out by fixing the parameters such as weight $\boldsymbol{w}$ and cluster center $\boldsymbol{V}$ as a constant. The membership value is considered as fixed constant in order to derive the optimal centers and weights to combine the kernel.

$$\frac{\partial J(w,U,V)}{\partial v_c} = -2\sum_{i=1}^{N} u_{ic}^m(\psi(x_i) - v_c) = 0$$

The above function is defined by using the derivative of $J(\boldsymbol{w},\boldsymbol{U},\boldsymbol{V})$ with respect to cluster center$\boldsymbol{v_c}$.

Hence, when $\boldsymbol{U}$ are given, the optimal $\boldsymbol{v_c}$ is the following closed form solution represented by the combination weights

$$v_c = \frac{\sum_{i=1}^{N} u_{ic}^m \psi(x_i)}{\sum_{i=1}^{N} u_{ic}^m} = \sum_{i=1}^{N} \hat{u}_{ic}\psi(x_i) \quad (4)$$

Where $\hat{u}_{ic} = \frac{u_{ic}^m}{\sum_{i=1}^{N} u_{ic}^m}$ is the normalized membership.

The cluster center in the kernel space cannot be evaluated directly because it may be either implicit or has an infinite dimensionality. If the cluster center in the kernel space is defined as a closed-form optimal solution then the finding the weights is possible for fixed membership and it is possible to evaluate the cluster centers directly by using the equation 4.

$$\begin{aligned} D_{ic}^2 &= (\psi(x_i) - v_c)^T(\psi(x_i) - v_c) \\ &= \psi(x_i)^T\psi(x_i) - 2\psi(x_i)^T\left(\sum_{j=1}^{N}\hat{u}_{jc}\,\psi(x_j)\right) \\ &+ \left(\sum_{j=1}^{N}\hat{u}_{jc}\,\psi(x_j)\right)^T\left(\sum_{j'=1}^{N}\hat{u}_{j'c}\,\psi(x_{j'})\right) \\ &= \sum_{k=1}^{M}\omega_k^2 k_k(X_i,X_j) \\ &- 2\sum_{j=1}^{N}\sum_{k=1}^{M}\hat{u}_{jc}\omega_k^2 k_k(X_i,X_j) \\ &+ \sum_{j=1}^{N}\sum_{j'=1}^{N}\sum_{k=1}^{M}\hat{u}_{jc}\,\hat{u}_{j'c}\omega_k^2 k_k(x_j,x_{j'}) \quad (5) \end{aligned}$$

Since memberships are fixed and kernel functions can be evaluated, equation (5) can be rearranged as

It is possible to rearrange the kernel function when the membership function is fixed.

$$D_{ic}^2 = \sum_{k=1}^{M}\alpha_{ick}\omega_k^2$$

where the coefficient α_{ick}can be written as

$$\begin{aligned} \alpha_{ick} = k_k(x_i,x_i) &- 2\sum_{j=1}^{N}\hat{u}_{jc}k_k(x_i,x_j) \\ &+ \sum_{j=1}^{N}\sum_{j'=1}^{N}\hat{u}_{jc}\hat{u}_{j'c}\,k_k(x_i,x_{j'}) \end{aligned}$$

It is to be observed that cluster centers from the evaluation is been eliminated. Thus, the objective function in equation (2) becomes

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

$$J(w,U)=\sum_{i=1}^{N}\sum_{c=1}^{C}u_{ic}^{m}\sum_{k=1}^{M}\alpha_{ick}\omega_k^2 \; subject\ to\ \omega_1 + \omega_2 + \cdots + \omega_M = 1\ and and\ \omega_k \geq 0\ \forall k\ and\ \sum_{c=1}^{C}u_{ic} = 1\ \forall i \quad and\ u_{ic}\geq 0\ \forall i,c$$

$$J(w)=\sum_{k=1}^{M}\beta_k\omega_k^2 \text{ subject } to\ \omega_1+\omega_2+\cdots+\omega_M = 1 and\ \omega_k \geq 0\ \forall k$$

where the coefficient β_k is

$$\beta_k=\sum_{i=1}^{N}\sum_{c=1}^{C}u_{ic}^{m}\alpha_{ick} \qquad (7)$$

This is a constrained optimization problem. By introducing a Lagrange multiplier, we have

$$J_\lambda(w,\lambda)=\sum_{k=1}^{M}\beta_k\omega_k^2-2\lambda\left(\sum_{k=1}^{M}\omega_k-1\right)$$

The partial derivatives are defined to be zero and ignore the constraints weights for forming the

$$\frac{\partial J_\lambda}{\partial \omega_k}=2\beta_k\omega_k-2\lambda=0$$

The solution for the aforementioned equation is $\omega_k=\frac{\lambda}{\beta_k}$

$$\sum_{k=1}^{M}\omega_k=\left(\frac{1}{\beta_1}+\frac{1}{\beta_2}+\cdots+\frac{1}{\beta_M}\right)\lambda=1$$

And

$$\lambda=\frac{1}{\frac{1}{\beta_1}+\frac{1}{\beta_2}+\cdots+\frac{1}{\beta_M}}$$

and the weight is the harmonic mean

$$\omega_k=\frac{\frac{1}{\beta_k}}{\frac{1}{\beta_1}+\frac{1}{\beta_2}+\cdots+\frac{1}{\beta_M}} \qquad (8)$$

In earlier days, the alternative optimizations of memberships and kernel combination weights are derived. But, the derivations are based only on the equality constraints and do not take into consideration the inequality constraint, i.e., the memberships and weights should not be negative. Since it is easy to verify that the derived memberships always satisfy $u_{ic}\geq 0\ and\ \sum_{i=1}^{N}u_{ic}>0\forall c$, it is showed that the solution of the combination weights also satisfies the non-negative constraint, i.e., $\omega k \geq 0$.It is initially indicated that $\beta k \geq 0\ for\ all\ k$. By definition, D_{ic}^2 should always be nonnegative for all weights, i.e., $\forall \omega k, D_{ic}^2=\sum_{i=1}^{N}\alpha_{ick}\,\omega_k^2\geq 0$. Thus, it is concluded that $\forall \omega k, \alpha_{ick}$ 0. Otherwise, if $\alpha_{ick'}<0$ for some k', we can let $\omega_{k'}=1\ and\ \omega_k=0$ if $k\neq k'$. However, this set of weight assignments means that $D_{ic}^2<0$, which contradicts its non negative property.

3.2.4 Algorithm 2 Mutiple Kernel Fuzzy C-Means (MKFC)

Given a set of N data points $X=\{x_i\}_{i=1}^{N}$, a set of kernel functions $\{K_k\}_{k=1}^{M}$, and the desired number of clusters C, output a membership matrix $U=\{u_{ic}\}_{i,c=1}^{N,C}$ and weights $\{w_k\}_{k=1}^{M}$ for the kernels.

1: **procedure** MKFC (Data $\boldsymbol{X}$, Number C, Kernels $\{K_k\}_{k=1}^{M}$)
2: Initialize membership matrix $U^{(0)}$
3: **repeat**
4: $\hat{u}_{ic}^{(t)}=\frac{u_{ic}^{(t)^m}}{\sum_{i=1}^{N}u_{ic}^{(t)^m}}$ calculate normalized memberships
Calculate coefficients by Equation (15)
5: **for**($i=1..N; c=1..\ C; k=1..M$) **do**
6: $\alpha_{ick}\leftarrow K_k(\boldsymbol{X_i},\boldsymbol{X_i})-2\sum_{J=1}^{N}\hat{u}_{ic}^{(t)}\,k_K(\boldsymbol{X_i},\boldsymbol{X_i})+\sum_{j=1}^{N}\sum_{j=1}^{N}\hat{u}_{ic}^{(t)}\,\hat{u}_{ic}^{(t)}\,k_K(\boldsymbol{X_i},\boldsymbol{X_i})$
7: **end for**
Calculate coefficient by Equation (7)
8: **for**(k=1..M) **do**
9: $\beta_k\leftarrow\sum_{i=1}^{N}\sum_{c=1}^{C}(u_{ic}^{(t)})^m\,\alpha_{ick}$
10: **end for**
Update weights by Equations (8)
11: **for**($k=1..M$) **do**
12: $w_k^{(t)}\leftarrow\frac{\frac{1}{\beta_k}}{\frac{1}{\beta_1}+\frac{1}{\beta_2}+\cdots+\frac{1}{\beta_M}}$
13: **end for**
Calculate distance by Equation(6)
14: **for**($i=1..N; c=1..C$) **do**
15: $D_{ic}^2\leftarrow\sum_{k=1}^{M}\alpha_{ick}\ (w_k^{(t)})^2$
16: **end for**
Update memberships by Equation (3)
17: **for**($i=1..N; c=1..C$) **do**
18: $u_{ic}^{(t)}\leftarrow\frac{1}{\sum_{c'}^{C}(\frac{D_{ic}^2}{D_{ic'}^2})^{\frac{1}{m-1}}}$
19: **end for**
20: **until** $||\boldsymbol{U}^{(t)}-\boldsymbol{U}^{(t-1)}||<\epsilon$
21: **return** $\boldsymbol{U}^{(t)},\{w_k^{(t)}\}_{k=1}^{M}$
22: **end procedure**

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Therefore, since both α_{ick} *and* u_{ic} are nonnegative, from (18), it is concluded that $\beta k \geq 0$. Finally, since ωk *'s* are harmonic means of nonnegative βk *'s* as shown in (19), they are also nonnegative. Thus, the obtained solution satisfies this constraint even though the nonnegative constraint is not taken into an account.

The objective function does not use the cluster center value because it is not potentially assessable. The constraints such as non negative values and unity constraints are satisfying by the random membership function which is initialized in proposed algorithm. The multiple kernel fuzzy c-means clustering is explained in algorithm 2 briefly. The process of subset is calculated in proposed algorithm till it reaches the value of change in membership matrix below threshold. The $O(N^2CM)$ is time complexity of MKFC algorithm based on the iteration excluding the construction of the kernel matrices.

3.3 Fuzzy Neural Network

The combination of the fuzzy logic and artificial neural network is used in the neural network is explained in [24, 25]. The FNN is one of the important topics in the research field because it is used in various applications.

The basic functions of neural network is training the input data, output data, parameter connection between the neurons which is adjusted through the repeated error corrections and these functions are mainly used to achieve the purpose of learning. The normal if-then rules in the network cannot be encoded directly. The only method is giving a large number of training data to the system. When the fuzzy system is compared with the neural network, the input values can be directly encoded in the fuzzy systems and the tolerance level is also high in fuzzy based system than the neural network [26].

3.3.1 Fuzzy System

Fuzzification is mainly used to conversion of the input data into the fuzzy sets which contains membership function. The membership functions used here is based on the trigonometric functions. The mechanism of fuzzy system is based on inference mechanism which simulates the decision model. The rules of the fuzzy system is explained below as R^j

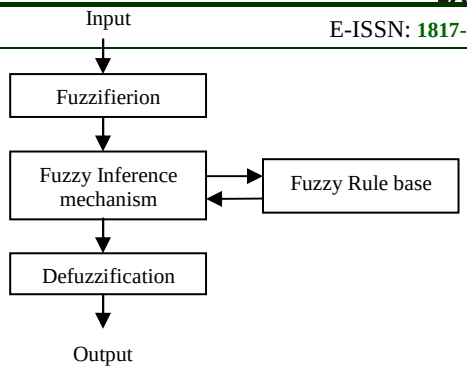

Figure 2: Fuzzy Inference Systems

$$R^j:$$

$$if\ x_1\ is\ A_1^j\ and\ x_1\ is\ A_2^j\ then\ x\ belongs\ to\ class\ y_j$$

$$with\ CF = CF_j \quad (2)$$

Where R^j denotes the j-th fuzzy rule, A_1^j and A_2^jindicates the fuzzy sets where $j = 1\ to\ n$ with n rules. $y_j \in y\{1,2,3 \ldots . M\}$ represents the output of the j-th fuzzy rule of M classes and CF_j denotes the reliability of the fuzzy rule R^j. The fuzzy inference outputs are obtained if Generalized Modus Ponens and Max-Min composition operation is applied in the equation. The obtained output is expressed as [22]:

$$\mu_y(y) = \max_j\{\mu_{y,j}\}\big|y_j = y$$

Where

$$\mu_{y,j} = \mu m_{A,j}(x).CF_j$$

$$\mu m_{A,j}(x) = \min\left\{\mu_{A_1^j}(x_1), \mu_{A_2^j}(x_1)\right\}$$

The defuzzification block is used to convert the fuzzified value into specific value. Usually the fuzzy inference output may be the fuzzy sets or specific values. If the output is fuzzy sets then the output is converted into the specific value by applying the median method and center area method. The output of the defuzzification is obtained by connecting the fuzzy value to the neuron directly.

3.3.2 The Single Neuron

The process of the calculating the value by processing the input signal along with the bond value using the neurons in the fuzzy system and the output obtained from the first layer is passed to the neurons in the next layer based on the condition of threshold value. The obtained values are trained in the neural network by using the back propagation algorithm based on the expectations of the feedback. As a result the bond values can be regulated by the network with the help of excitation

Journal of Theoretical and Applied Information Technology
10th March 2014. Vol. 61 No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

from external environment. The fuzzy rules can be simplified by using the learning process in the fuzzy neural network as aforementioned.

The fuzzy system is constructed based on the requirement for memory and computation time in a real-time system; in this study the single neuron is used to construct the real time system. The ability to map nonlinear function, in the fuzzy system is unfeasible in the real time system and this problem can be solved after the front end operation. The p external input is accepted and it is mainly used for calculating the output along with p bond value and transfer function$y = \varphi(v)$. The formula for calculating the output based on the classification inference probability is

$$v = \sum_{i=0}^{p} w_i u_i$$

$$y = \varphi(v) = \frac{\alpha_2}{1 + e^{-\alpha_1 . v}}$$

Where the shape of the function can be controlled by using α_1 and the scale size can be adjusted by usingα_2. Fig. 3 shows the combination of the fuzzy system and the single neuron.

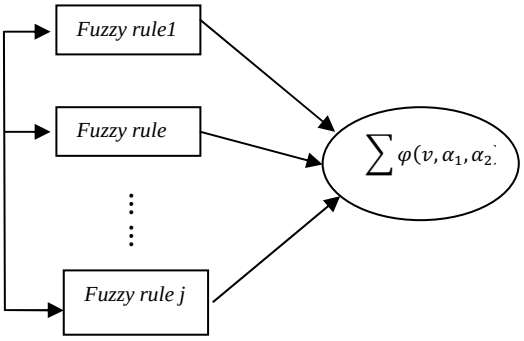

Figure 3: The Artificial Neural Network combined with single neuron

3.4 Fuzzy Aggregation Module

The collection of FNN outcomes is the main goal of the fuzzy aggregation module which is mainly used to trim down the detected errors in the subset of TR_i in every module of FNN. The steps of FNN process to learn the errors are explained below

Step 1: The training set TR data is considered as an input to the every trained FNN_i and get the outputs:

$$Y_j^{TR} = \left[y_{j1}^{TR}, y_{j2}^{TR}, \dots, y_{jk}^{TR}\right], j = 1,2 \dots, n$$

Where n represents the number of training set $TR, and\ y_{jk}^{TR}$ represents the output of FNN_k.

Step 2: The input to the new FNN is represented as a

$$Y_{input} = [Y_1^{TR}, U_1^{TR}, \dots \dots, Y_n^{TR}, U_n^{TR}]$$

where U_n^{TR} represents the membership grade of TR_n belonging to C^{TR}.

Step 3: The output of the new FNN is obtained by training the FNN using the Y_{input} and whole training set TR's class label.

The process of both FNN and fuzzy aggregation module is same as the above mentioned process in testing stage. The membership value is calculated by using the cluster centerC^{TR}, input x_i^{TS} and formula is defined as

$$u_{ji}^{TS} = \frac{1}{\sum_{p=1}^{k} \left(\frac{\left\| x_i^{TS} - c_j^{TR} \right\|}{\left\| x_i^{TS} - c_p^{TR} \right\|} \right)^{\frac{2}{m-1}}}$$

The Y_{output}^{TS} can be obtained by using FNN module and fuzzy aggregation module.

4. EXPERIMENTS AND RESULTS

4.1 Data Preparation

In [27] the system is experimented using KDD CUP 1999 dataset and it is worn by MIT Lincoln Laboratory's DARPA intrusion detection valuation program. The five million training dataset, two million testing dataset are used in the proposed dataset based on the 41 features of the dataset is reclaimed from the each connection and connection record status is specified as attack type. The reclaimed features in the connection are sort out into four types such as 1) the TCP connection features such as connection period, type of protocol, network service are included in the intrinsic connection. 2) The payload of the TCP connections is accessed by using the content features within the connection. 3) The connection feature is used to calculate the statistics related to the protocol behavior service and the connection between the host and destination is recognized 4) the connection feature is used to examine the connection between the host and destination for past two seconds which has the same service as current connection.

The intrusion detection system is used to detect the attacks and the attacks are categorized into four types. They are explained as 1) DoS Denial of service attack 2) Probe 3) Remote to Local (R2L)

Journal of Theoretical and Applied Information Technology
10th March 2014. Vol. 61 No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

4) User to Root (U2R). These types' of attacks can be found easily by using the proposed intrusion detection system.

The dataset is preprocessed to reduce the size of the dataset by using the method called random selection [28]. The 19525 records are selected as an input to the proposed system. The selected input records consist of normal records and records with aforementioned attacks also. These dataset is used as a training dataset and KDD testing dataset is used for testing purpose. Some of the specific types of attacks based records are specified in the testing dataset which is not present in the training dataset.

4.2 Evaluation Criteria

The parameters like true positives, true negatives, false positives, and false negatives are mostly used for assess the intrusion detection system [29]. The particular attack can detect by using the IDS exactly is called as true positives. The normal condition can be found without the error is called true negatives. The false attack can be found by using the IDS are called false positives. The false positive attack happened is based on the loose recognition, limitation on detection or particular environmental factor. Some type of attacks cannot be found by IDS because the pattern or rule for attack is does not exist is called false negatives. Some types of attacks in the dataset are not in large amount so it cannot be used for evaluating the proposed system and it will be biased. The evaluating parameters such as precision recall F-value are used in this study to evaluate the proposed system. The formula's are defined as

$$precision = \frac{TP}{TP + FP}$$

$$recall = \frac{TP}{TP + FN}$$

$$F - value = \frac{(1 + \beta^2) * recall * precision}{\beta^2 * (recall + precision)}$$

where TP, FP, and FN represents the number of true positives, false positives, and false negatives, respectively, and β represents to the relative importance of precision versus recall and is usually set to 1.

Some of the disadvantages of the artificial neural network is sometimes it is unstable because it is commonly failed to converge the local minimum and failed to train the dataset. It will affect the performance of the proposed system to overcome this fuzzy neural network is proposed. The percentage of the evaluation criteria is also measured by using the parameters ie., successful training is measured by using the detection stability.

$$\text{percentage of training successfully} = \frac{\text{the number of training successfully}}{\text{the number of training}}$$

5. RESULTS AND DISCUSSIONS

In the experiments results, the connection features are mentioned in the form of vector. The retrieved features are continuous, nominal and discrete. The algorithms based on the clustering and classification is used to find the continuous values. If the feature is nominal means then it is converted into continuous and it is given as input to the system. In MKFC clustering module, the training dataset is converted into subset based on the fuzzy. The layer used in the fuzzy neural network is 41 nodes because it is mainly based on the retrieved features and 5 nodes are used in the fuzzy aggregation module based on the types of attacks.

The formula for predicting the number of nodes in the neural network is

$$\sqrt{I + O} + \alpha(\alpha = 1\text{–}10),$$

I Represent the number of input node,

O Represents the number of output node,

a Random number

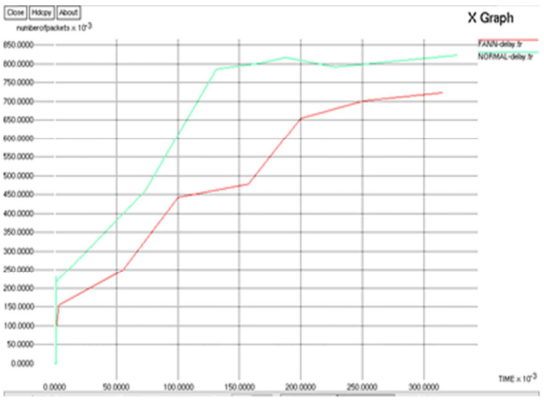

Figure 4: Comparison of delay

The comparison of delay between the FC-ANN and MKFCM-FNN is represented in the graph in fig 4. The delay gets increased as the number of packets increases. It is observed from the graph that the proposed MKFCM-FNN outperforms the existing approach in terms of delay.

Journal of Theoretical and Applied Information Technology
10th March 2014. Vol. 61 No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

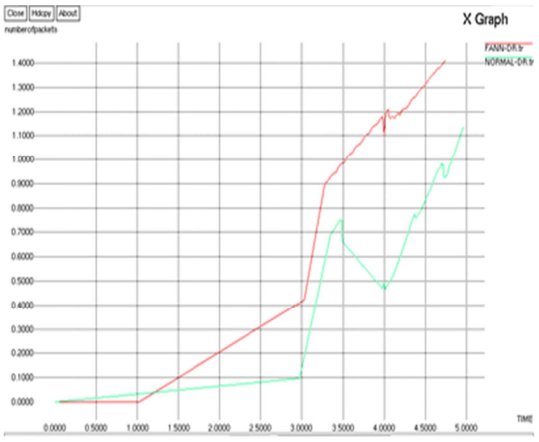

Figure 5: Comparison of Detection Ratio

The comparison of detection ratio between the FC-ANN and MKFCM-FNN is represented in the graph in fig 5. The detection ratio of the proposed MKFCM-FNN is compared with the existing system detection ratio. The system performance is measured based on the detection ratio and the proposed approach outperforms the existing system.

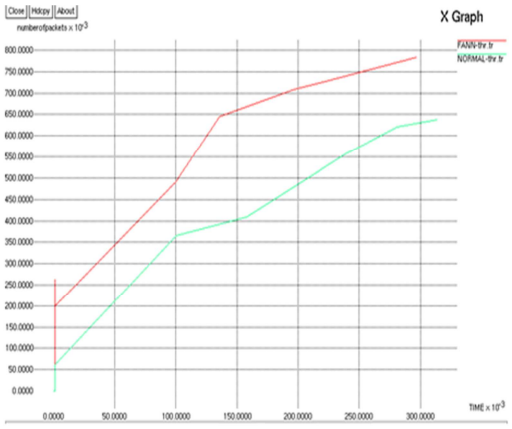

Figure 6: Comparison of threshold

Figure 6 shows the Threshold comparison of the FC-ANN and MKFCM-FNN. It is clearly observed from the graph that the proposed MKFCM-FNN based IDS approach outperforms the existing FC-ANN approach.

The experiments is conducted and the simulated results are compared with the other typed of intrusion detection system. The parameters are taken here are measurement of precision, recall and f-value of the systems.

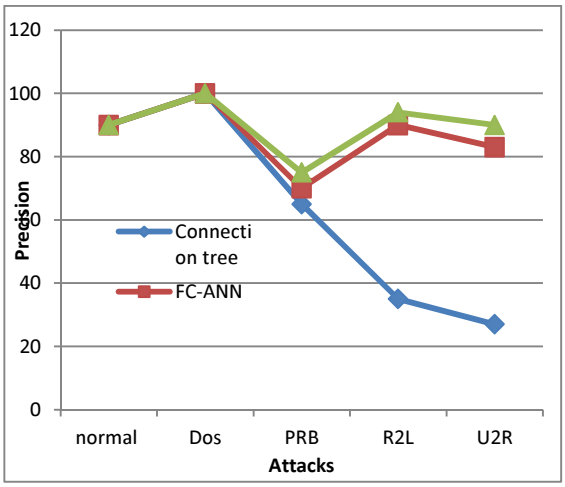

Figure 7: Precision Comparison (%)

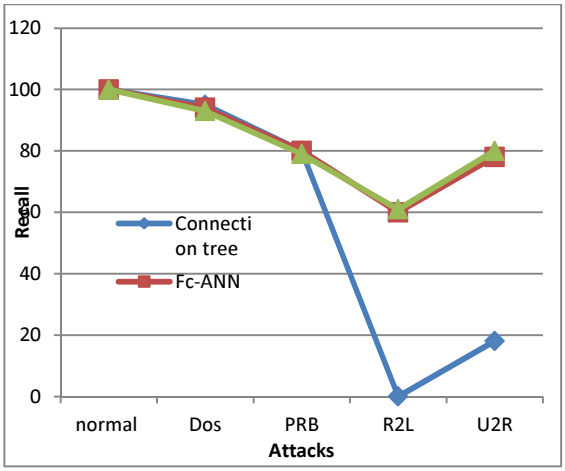

Figure 8: Recall (%) of different methods.

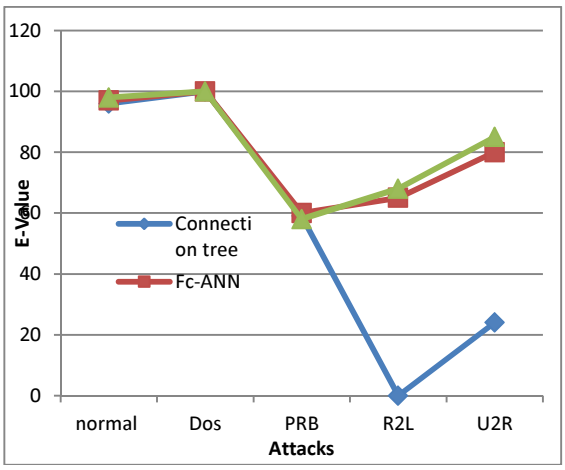

Figure 9: F-value (%) of different methods

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

6. CONCLUSION AND FUTURE WORKS

Intrusion detection system is mainly used to detect the attacks in the system. In recent years many research has been conducted based on the intrusion detection system and many real time application is also created based on the IDS systems. In this paper, fuzzy neural network and multiple kernel fuzzy c-means clustering based system is proposed for intrusion detection system. The problem in the existing intrusion detection system is overcome in this proposed system. The proposed system is compared with the existing system and it is evaluated by using the measures such as precision, recall and f-value and it gives better results.

REFRENCES:

[1] K.-C. Lan, A. Hussain, and D. Dutta, "Effect of Malicious Traffic on The Network," *Proc. Passive and Active Measurement Wksp. (PAM)*, San Diego, CA, Apr. 2003.

[2] S.-X. Wu and W. Banzhaf, "The Use of Computational Intelligence in Intrusion Detection Systems: A Review," *Elsevier Applied Soft Computing*, vol. 10, issue 1, Jan. 2010, pp. 1–35.

[3] H. T. Elshoush and I. M. Osman, "Reducing False Positives through Fuzzy Alert Correlation in Collaborative Intelligent Intrusion Detection Systems — A Review," *Prof. IEEE Int'l. Conf. Fuzzy Systems*, July 2000, pp. 1–8.

[4] Kabiri P, Ghorbani AA. Research in intrusion detection and response – a survey. International Journal of Network Security 2005;1(2):84–102.

[5] Sobh TS. Wired and wireless intrusion detection system: classifications, good characteristics and state-of-the-art. Computer Standards & Interfaces 2006;28:670–94.

[6] Denning ED. An intrusion-detection model. IEEE Transactions on Software Engineering 1987;13(2):222–32.

[7] Staniford-Chen S., Tung B., Porrar P., Kahn C., Schnackenberg D., Feiertag R., et al. The common intrusion detection frameworkdata formats. 1998. Internet draft 'draft-staniford-cidf-dataformats- 00.txt'.

[8] Wang, W., Behera, S. R., Wong, J., Helmer, G., Honavar, V., Miller, L., Lutz, R., & Slagel, M. (2006). Towards the automatic generation of mobile agents for distributed intrusion detection system. *Journal of Systems and Software*, *79*, 1-14. Retrieved from www.elsevier.com/ locate/jss

[9] Anderson, James P., "Computer Security Threat Monitoring and Surveillance",Fort Washington, Pa., 1980.

[10] Endorf, C., Schultz, E., & Mellander, J. (2004). Intrusion detection and prevention. California: McGraw-Hill.

[11] Silva, L. D. S., Santos, A. C., Mancilha, T. D., Silva, J. D., & Montes, A. (2008). Detecting attack signatures in the real network traffic with ANNIDA. Expert Systems with Applications, 34(4), 2326–2333.

[12] Patcha, A., & Park, J. M. (2007). An overview of anomaly detection techniques: Existing solutions and latest technological trends. Computer Networks, 51(12), 3448–3470.

[13] Manikopoulos, C., & Papavassiliou, S. (2002). Network intrusion and fault detection: A statistical anomaly approach. IEEE Communications Magazine, 40(10), 76–82.

[14] Dokas, P., Ertoz, L., Lazarevic, A., Srivastava, J., & Tan, P. N. (2002). Data mining for network intrusion detection. Proceeding of NGDM, 21–30.

[15] Ryan, J., Lin, M., & Miikkulainen, R. (1998). Intrusion detection with neural networks. Advances in neural information processing systems (Vol. 10). Cambridge, MA: Springer.

[16] G. P. Spathoulas and S. K. Katsikas, "Using a Fuzzy Inference System to Reduce False Positives in Intrusion Detection," *Proc. 16th Int'l. Conf. Systems, Signals and Image Processing*, June 2009.

[17] M. Sourour, B. Adel, and A. Tarek, "Environmental Awareness Intrusion Detection and Prevention System toward Reducing False Positives and False Negatives," Proc. IEEE Symp. Computational Intelligence in Cyber Security, Apr. 2009.

[18] I.-W. Chen *et al.*, "Extracting Attack Sessions from Real Traffic with Intrusion Prevention Systems," *Proc. IEEE ICC*, June 2009.

[19] S.-H. Wang, "Extracting, Classifying and Anonymizing Packet Traces with Case Studies on False Positives/Negatives Assessment," M.S. thesis, Dept. Comp. Sci., Nat'l. Chiao Tung Univ., Taiwan, 2010.

[20] Y.-D. Lin *et al.*, "On Campus Beta Site: Architecture Designs, Operational Experience, and Top Product Defects," *IEEE Commun. Mag.*, vol. 48, no. 12, Dec. 2010, pp. 83–91.

[21] M. Roesch. Snort - Lightweight Intrusion Detection for Networks. In *LISA '99: Proceedings of the 13th USENIX conference*

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

on System administration, pages 229–238, Berkeley, CA, USA, 1999.

[22] J. Shawe-Taylor and N. Cristianini, *Kernel Methods for Pattern Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[23] Kosko, Bart (1992). Neural Networks and Fuzzy Systems: A Dynamical Systems Approach to Machine Intelligence. Englewood Cliffs, NJ: Prentice Hall. ISBN 0-13-611435-0.

[24] J. Lin, W. J. Hwang, and R. J. Wai, 1999, "A supervisory fuzzy neural network control system for tracking periodic inputs," IEEE Trans. Fuzzy Systems, Volume 7, No.1, pp. 41-52.

[25] Y. C. Chen and C. C. Teng, 1995, "A model reference control structure using a fuzzy neural network," Fuzzy Sets and Systems, Volume 73, pp.291-312

[26] *V. Paxson. Bro: A System for Detecting Network Intruders* in *Real-Time. In Computer Networks, volume 31 (23–24), pages 2435–2463, 1999.*

[27] Bezdek, J. (1974). Fuzzy mathematics in pattern classification. Ph.D. thesis. Ithaca, NY: Cornell University.

[28] T.D. Pham and X. Liu, Neural Networks for Identification, Prediction and Control, Great Britain, 1995.

[29] G.A. Darbellay and M. Slama, Forecasting the short-term demand for electricity. Do neural networks stand a better chance? International Journal of Forecasting, 16:71-83, 2000.

[30] KDD CUP 1999 dataset (1999). <http://archive.ics.uci.edu/ml/datasets/KDD+Cup+ 1999+Data> (accessed March 2009).

[31] Beghdad, R. (2008). Critical study of neural networks in detecting intrusions. Computers and Security, 27(5-6), 168–175.

[32] Axelsson, S. (2003). The base-rate fallacy and the difficulty of intrusion detection. ACM Transaction on Information and System Security, 3, 186–205.

[33] Witten, I. H., & Frank, E. (2005). Data mining: Practical machine learning tools and techniques. Boston: Morgan Kaufmann Publishers