ISSN : 2347 - 8446 (Online)
ISSN : 2347 - 9817 (Print)

A Hybrid Intrusion Detection System for WSN

[I]Alaa Eissa, [II]Hazem M. El Bakry, [III]Samir M. Zaid
[I,II]Dept. Information Systems, Faculty of Computer Science and Information Systems, Mansoura University, Mansoura, Egypt
[III]Dept. of Geology, Faculty of Sciences, Zagazig University, Egypt

Abstract
Wireless sensor networks is a collection of homogenous and limited resource sensors that deployed with a dense manner in an area of interest and communicate with each other to accomplish its tasks. The wireless nature of this networks make it a victim to intrusions, so it is important to provide security mechanism such as intrusion detection systems. Intrusion detection is the process of observing and analyzing the events that done in a computer system or network to find an intrusion which violate computer or network security. This paper provide an intrusion detection system for wireless sensor networks that overcome the previous systems limitations and provide high detection rate with low false alarm. Our intrusion detection system composed of two phases, in the first phase data set is clustered to two clusters normal cluster and attack cluster using PFCM clustering algorithm. Then at second phase clustered data is moved to multi-layer perceptron neural network that classify it to normal and attacks types DOS, Probe, U2R, R2L using back propagation training algorithm as it provide high performance.The results of implementation show that our system provide high detection rate with low false alarm.

Keyword
Artificial Neural Network, Clustering, Data Mining, Intrusion Detection, Wireless Sensor Networks

I. Introduction

Wireless sensor networks (WSNs) is an important field of research as most researchers focus their attention on this field and it is used in different areas such as industrial applications, business applications and military applications. This network contain relatively inexpensive and well distributed sensor nodes, one or more powerfull nodes used to collect data from sensors, this nodes called sink nodes. All nodes composed from the following units (wireless radio transceiver that transmit and receive communications, a tiny microcontroller, a power supply and multi-type sensors). Wireless sensor networks vary from traditional wireless communication networks, as WSNs impose special and unique characteristics, for example, intensively deployment of nodes, unreliability properties, restrictions in energy, computation, and storage, so significant challenges presented to the developing and improving WSN applications.

Wireless sensor networks is unsafe environment as data collected by sensor nodes is often unreliable that because collected data may be affected by noise and error.

Intrusions can be done in WSN by several reasons such as errors, malfunction and attacks. Intrusions have three classifications (data Intrusions, network Intrusions and node Intrusions) [11-14]. It is not easy to design an intrusion detection system (IDS) for WSNs because of the special requirements of WSN. Intrusion detection approaches used in wired network and other types of wireless network can not be used in WSNs due to WSN characterstics [1]. In a traditional wired network, the detection system detects attacks based on the signature in a centralized technique as data are sent to a high resource server to be analyzed [2].

Intrusion in wireless sensor networks can be detected using two different approaches which is anomaly detection and signature based detection. signature based techniques concentrate on features of known attacks so it have high detection rate and low false alarms but it have a disadvantage which is inability to detect new attacks. On the contrary anomaly detection techniques concentrate on learning the feature of normal traffic not on attack traffic and the behavior that violate the normal behavior is announced as attack, this techniques have the ability to detect new attacks that never seen before but it suffer from high false alarms.

Another point of view can classify the detection schema into hierarchal and flat depending on their architectures. In a hierarchical architectures all sensor nodes not have the same role and capabilities, there are some nodes which have a special roles, when nodes are clustered there are a node that selected to be a cluster head, this node must has a special characteristics as power, processing and transmission capabilities. In contrast, in flat architectures all sensor nodes have the same role and capabilities.

Wireless sensor networks are very sensitive network and it is susceptible to attacks because of their sensitivity nature, so they require intrusion detection system to monitor traffic for anomalies at multiple concentration points. But this security requirements need special characteristics as high energy consumption, large memory capabilities and high bandwidth for transmission that not available in wireless sensor networks as it is resource constrained. Intrusion detection system that applied to wireless sensor network must have some characteristics as fully distributed and inexpensive communication, energy, and memory requirements. This systems must understand the features of attacks to have the ability to detect this intrusions [3].

II. Literature review

A lot of researcher introduced intrusion detection systems that used for particular applications. Although there are some of them used for general purpose. A general system is introduced in [4], it is used to discover and detect localization anomalies that done by adversaries. It is anomaly intrusion detection. Another intrusion detection system is introduced in [5], the system using an algorithm to generate the appropriate signatures of sensor in automatic manner. It present an intrusion detection approach which uses network topology information and sensors placement to determine what is announced as malicious in the network. An intrusion detection system presented in [6] that designed to improve the intrusion tolerance against base station isolation, it provide secure various path routing to multiple destination base stations. It also can disguise base station location by using anti-traffic analysis strategies. In [7] an intrusion detection approach is introduced, this approach use two algorithms one of them for

identifying erroneous sensor and the second algorithm for fault-tolerant event boundary detection. These algorithms providing two characteristics for wireless sensor networks as localizing and scalability. Da Silva et al: [8] introduce a hybrid anomaly detection system. In this system some special nodes called monitoring nodes are reliable to observe their neighbors to detect any abnormal action and so detect the anomalies. These nodes pay attention to messages in their radio scope and store certain message handle that may be valuable to the guideline application stage. At that point, they attempt to recognize a few assaults, similar to message delay, redundancy, information adjustment, black hole and selective forwarding. It is finished up from the paper that the buffer size to store the checked messages is a vital variable that enormously impacts the false positives number. Given the limited memory accessible in motes, all things considered the detection efficiency is kept to lower levels. Onat and Miri [9], introduce comparable intrusion detection system where every node has a constant-size buffer to store some information about the messages that received it such as packets arrival time and received power. The packet is donated as anomalous if its power is not within certain range or it arrived at different time. This system raises an alert of intrusion if the detected anomalies packets rate is over than a given threshold. Along these lines the authors assert that it is thinkable for a node to successfully recognize an intruder imitating a legal neighbor.
A lot of previous researches provide IDSs that used neural networks method, this show the importance of neural network technique that used for classification problems, Mohammed Sammany introduce hybrid intrusion detection system which can differentiate attack records and its type, he used a mix of K-means clustering algorithm and Multi-Layer Perceptron (MLP) Neural Network for classification, this system used MLP neural network that aimed to solve a multi-class problem. In test phase system provide accuracy rate 93.43% with low false alarm rate which considered a big challenge in intrusion detection system research. He introduced a system that provide high rate for detecting novel attacks with low false alert. Fuzzy logic and neural network can also combined together to build hybrid intrusion detection system with high detection rate and low false alarms. This system is to take advantage of each technique and overcome the disadvantages, and improving the accuracy of intrusion detection.

III. Research Method

Monitoring data have an overlap, so it is a great challenge to distinguish between normal and abnormal behavior in computer networks. We use fuzzy clustering to reduce the problem of interfere between typical and strange conduct. Thus our system intended to produce results with low false alarms and high detection rate. The research method in this paper is to use KDD99 Cup dataset as an input to modified Fuzzy C-Means clustering algorithm which is Possibilistic Fuzzy C-Means clustering algorithm (PFCM) for clustering data to two partitions one cluster is normal and the second cluster is attack cluster, then Multi-layer Perceptron neural networks (MLP) is used for classification to improve the performance of detection and identifying types of attack which are DOS (Denial of Service), Probe, U2R (User to Root), R2L. As shown in the following block diagram:

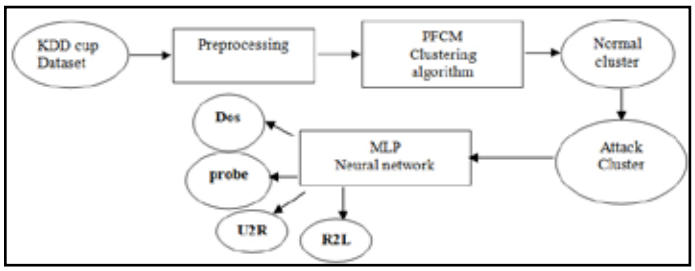

Fig. 1: Block diagram of proposed System

A. Possibilistic Fuzzy C-Means Clustering Algorithm

Fuzzy-possibilistic c-means (FPCM) algorithm is introduced in 1997, this algorithm is used for clustering labeled and unlabeled data, when it used for clustering unlabeled data it can generate membership and typicality values. The summation of typicalities over all data points belong to a cluster is one because FPCM algorithm constrains the typicality values. Large data sets have unrealistic typicality values because of row sum constraint. We used an algorithm called possibilistic-fuzzy c-means (PFCM). Memberships and possibilities can be produced simultaneously, along with the usual point prototypes or cluster centers for each cluster. This algorithm is a combination of fuzzy c-means (FCM) and possibilistic c-means (PCM) which avoids the limitations of fuzzy c-means, possibilistic c-means and fuzzy-possibilistic c-means. As each of previous algorithms have it's limitations for example FCM have a disadvantage which is noise sensitivity but it avoided by PFCM, PCM suffer from the coincident clusters problem which be avoided by PFCM and this algorithm overcome the row sum constraints problem of FPCM. Because of derived advantages of PFCM which is less sensitivity to outliers and it overcome the coincident clusters problem, this algorithm considered a strong candidate for fuzzy rule-based system identification. PFCM lead to optimize the following objective function [10]:

$$\min\left\{ J_{m,\eta}(U,T,V;X)=\sum_{k=1}^{n}\sum_{i=1}^{c}\left(au_{ik}^{m}+bt_{ik}^{\eta}\right)^{*}\left\|x_k-v_i\right\|_A^2+\sum_{i=1}^{c}\gamma_i\sum_{k=1}^{n}\left(1-t_{ik}\right)^{\eta}\right\}$$

subject to the constraints $\sum_{i=1}^{c}u_{ik}=1\forall k$ and $0\le u_{ik},t_{ik}\le 1$. Here a>0 , b>0 , m>1, η >1 and $j_{m,\eta}$ is the objective function. U is the partition matrix. T is the typicality matrix. V is a vector of cluster centers, X is a set of all data points, x represents a data point, n is the number of data points and c is the number of cluster centers which are described by s coordinates. $\|x_k-v_i\|_A$ is any norm used to calculate the distance between ith cluster center and kth data set(also represented by DikA). The constants a is fuzzy membership relative importance and b is the typicality values in the objective function. PFCM and FCM have the same meaning of membership and PFCM has the same typicality interpretation as in PCM. The importance of membership and importance of typicality are inversely proportional with the same amount. The optimal typicality values depend on the magnitude of b. So by constraining a+b=1, we lose modeling flexibility. PFCM algorithm is shown below:

- **Step1**: initialize U= [u_{ik}] matrix, $U^{(0)}$
- **Step2**: At K step: calculate the center vectors $C^{(k)}$ =[v_i] with $U^{(k)}$

ISSN : 2347 - 8446 (Online)
ISSN : 2347 - 9817 (Print)

$$v_i = \frac{\sum_{k=1}^{n}\left(au_{ik}^{m} + bt_{ik}^{\eta}\right)x^{k}}{\sum_{k=1}^{n}\left(au_{ik}^{m} + bt_{ik}^{\eta}\right)} \qquad 1 \le i \le c$$

- **Step3**: update $U^{(k)}$, $U^{(k+1)}$

$$u_{ik} = \left(\sum_{j=1}^{c}\left(\frac{D_{ikA}}{D_{ikA}}\right)^{2/(m-1)}\right)^{-1}$$

- **Step4**: if $\left\|U^{(k)} - U^{(k+1)}\right\| < \varepsilon$, then stop; otherwise return to step 2.

Where $$t_{ik} = \frac{1}{1 + \left(\frac{b}{\gamma_i} D_{ikA}^{2}\right)^{1/(\eta-1)}} \qquad 1 \le i \le c,\ 1 \le k \le n$$

B. MLP neural network

MLP is a multiplayer feed-forward neural network that contain an input layer, and an output layer and between this two layers there are one or multiple hidden layers. The output layer supplies the response of the network to the activation patterns applied to the input layer. Assigning the input pattern to one of patterns in output layer is the objective of MLP neural network.

The structure of artificial neural network is defined as input layer consist of 41 neuron because we use 41 feature, number of neurons in output layer is 4 neuron as we classify data set to 4 categories (Dos, Probe, U2R, R2L), we used one hidden layer with 22 hidden neurons that chosen based on some experiments, we choose initial random value for hidden neuron for example 8 hidden neuron and it incremented by 2. Our results show that when the number of hidden neuron is equal to 22 hidden neuron, the classification rate was 99.73%. We choose back propagation training algorithm for training the network as it is proved that the best algorithm that used for training is back propagation which has some advantages of taking less time with a small epoch number, and high precision. Back propagation algorithm is shown below:

- **Step1**: initializing weights with a small values
- **Step2**: provide the input (X) and specify the desired output (d)
- **Step3**: calculate the actual output (Y)
- **Step4**: Adjust weights through $W_{ij}(t+1) = W_j(t) + \mu\delta x_i$

  where μ is learning rate and δ is error term for node j, if j is output node then $\delta_j = y_j(1-y_j)(d_j - y_j)$, if j is a hidden node then $\delta_j = x_j(1-x_j)\sum_k \delta_k w_k$ where k is all nodes in layers above node j
- **Step5**: If the MSE is greater than predefined value then go to step 2 else finish.

IV. Results and Analysis

Training data have an important affect on neural network performance. Data collecting is a critical phase in developing any system. We used the KDD CUP 99 CSV file format data set as the input to our system. We download KDD Cup 99 data set from its home page. The data is downloaded in text format that not suitable for using as an input for our IDS, so we will prepare it for using. The first step is to convert this data set from text format to comma separated values. Kdd99 data set have some features which have all forms of continuous, discrete, and symbolic variables, Symbolic variables must be converted to numeric form with the goal that it can be given as inputs to our developed system, we can doing this step using Weka 3.6.2 program and apply nominal to binary filter. We used all 41 features to implement our system. After the preprocessing phase of KDD99 chosen datasets, dataset which contains 41 features is clustered using Possibilistic Fuzzy C-Means clustering algorithm (PFCM) to two clusters one cluster is normal and second cluster is attack. We use two data sets to implement our system, the first data set contain 22,133 records (998 normal records & 21135 attack records), and second data set contain 65,325 records (11739 normal records & 53586 attack records). The principal phase of the PFCM algorithm is to give an initial value to the input variable, the input vector comprises of 41 features as showed above, there are two clusters which are normal and attack, and the cluster center is computed by taking the means for all feature from arbitrary records in dataset. The results after applying PFCM to two datasets is shown in table 1, using first data set, after iteration 1 the number of records in normal cluster is 1720 and records in attack cluster is 20413, in iteration 2 the number of records in normal cluster is 1035 and records in attack cluster is 21098, in iteration 3 the number of records in normal cluster is 1007 and records in attack cluster is 21126, in iteration 4 the number of records in normal cluster is 1004 and records in attack cluster is 21129 and in iteration 5,6 it is the same as iteration 4 so we will stop. Table 1 show that when we using the second data set, it provide higher classification rate at iteration 4 also, so our algorithm will stop at iteration 4.

Table 1: result of the clustering using PFCM algorithm

	Input Data	iteration 1	iteration 2	iteration 3	Iteration 4	iteration 5	iteration 6
First data set	**Normal 998**	1720	1035	1007	1004	1004	1004
	Attack 21135	20413	21098	21126	21129	21129	21129
	Normal Classification	%58.023	%95.151	%99.542	%99.914	%99.914	%99.914
	Attack Classification	%96.582	%98.892	%99.911	%99.973	%99.973	%99.973
Second data set	**Normal 11739**	19743	12324	11790	11749	11749	11749
	Attack 53586	45852	53001	53535	53576	53576	53576
	Normal Classification	%59.45	%95.25	%99.56	%99.92	%99.92	%99.92
	Attack Classification	%85.56	%98.91	%99.90	%99.98	%99.98	%99.98

We calculated True Positive (TP) that mean if it is attack and detection system is attack (number of records classified as attacks divided by total number of attacks), True Negative (TN) mean if it is normal and detection system is normal (number of records classified as normal divided by total number of normal), False Negative (FN) means if it is attack and detection system is normal (number of attack instances that were classified as normal divided by the total number attacks) and False Positive (FP) means if it is normal and detect system is attack (number of normal instances that were classified as intrusions divided by the total number of normal instances. After compute the all previous parameters, we will compute the Detection rate, Accuracy and False alarm from the following equations:

$$DetectionRate = \frac{TP}{TP + FP} \qquad Acuracy = \frac{TP + TN}{TP + TN + FP + FN} \qquad FalseAlarm = \frac{FP}{FP + TN}$$

The following table showing the values of (TPR- TNR- FPR- FNR):

Table 2: Our IDS results (TPR- TNR- FPR- FNR)

	True Positive (TPR)	True Negative (TNR)	False Positive (FPR)	False Negative (FNR)
First data set	99.973%	99.914%	0.086%	0.027%
Second data set	99.98%	99.92%	0.08%	0.02%

Table 3: Comparison (FPR&TPR) with previous systems

Classifier	**False Positive (FPR)**	**True Positive (TPR)**
Decision Tree	4.2%	95.8%
Support Vector Machine (SVM)	10.3%	89.7%
K_Nearest Neighbor(K_NN)	6.5%	93.5%
K_Meanse Clustering	7.0%	93.5%
Our IDS	0.08%	99.98%

From the previous values, we can calculate the detection rate, accuracy and false alarm, the detection rate is 99.98%, accuracy is 99.95% and false alarm is 0.08. This result show that our IDS provide high detection rate with low false alarms and this is the purpose of our intrusion detection system.

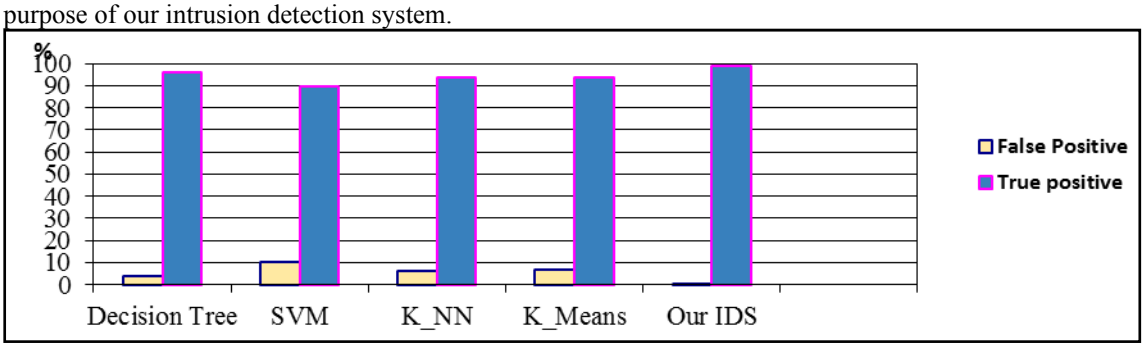

Fig. 2: Comparison False positive & True Positive

PFCM clustering stage are used to clustering data set into two clusters to minimize the complexity, the second stage of neural network is used to improve the performance of detection rate and utilized to classify attack records to it's four sorts. Multi-layer feed forward networks (MLP) is used. After network training, network now can be tested using two test data sets, the first data set contain 21129 attack records, second data set contain 53576 attack records. After loading test data and implement our IDS the results shown in tables below:

Table 4: Testing using first data set

	Input test data	Corrected identified data	Classification rate
Dos	14024	13903	99.13%
Probe	6371	6327	99.30%
U2R	104	79	75.9%
R2L	630	276	43.8%

ISSN : 2347 - 8446 (Online)
ISSN : 2347 - 9817 (Print)

Table 5: Testing using second data set

	Input test data	Corrected identified data	Classification rate
Dos	47689	47288	99.16%
Probe	4127	4102	99.39%
U2R	1719	1289	75%
R2L	41	0	0%

The above results showing that our IDS provide high classification rate for Dos records, Probe records and U2R classification rate is acceptable to some extent but it classify R2L records with low percentage which can be improved in future works.
When we compare our IDS with another previous systems, results show that our system is more effective and provide high detection rate

Table 6: comparison of classification rate with previous systems

	Our IDS	SOM	Improved SOM
Dos	99.145%	95.66%	96.74%
Probe	99.345%	83.5%	91.0%
U2R	75.45%	10.0%	43.33%
R2L	21.9%	5.45%	20.0%

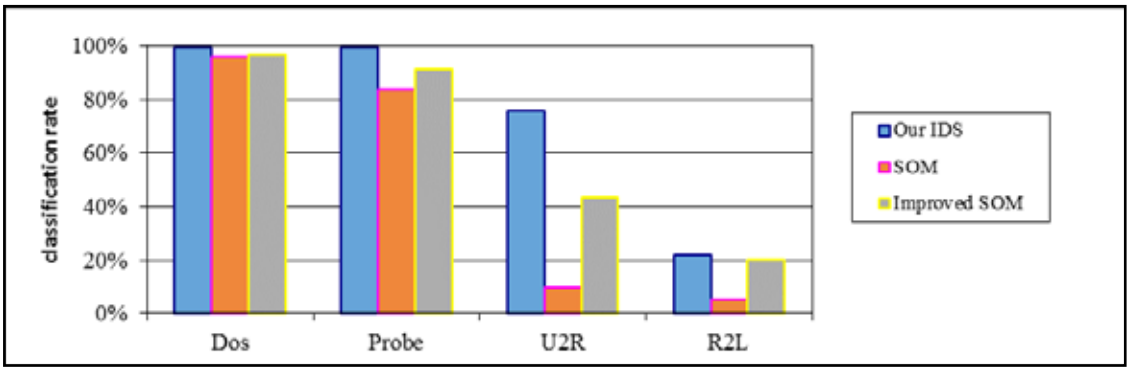

Fig. 3: A comparison between classification rates

V. Conclusion

An intrusion detection system that composed of two phases has been presented. In the first phase data set is clustered to two clusters normal cluster and attack cluster, then attack cluster is moved to multi-layer perceptron neural network that classify it to attacks types DOS (Denial of Service), Probe, U2R (User to Root), R2L using back propagation training algorithm as it provide high performance. We built this system using Java programming and Weka tool [Neural Network visualizer] and to run our system you need to install jdk1.8 and above and NetBeans IDE 8.0.2. Two data sets from KDD data set have been used to test our system which contain attack and normal records. Data set must be preprocessed before clustering. After preprocessing data set is moved into PFCM clustering algorithm which is a modification to FCM, the clustering algorithm divide it into to clusters normal cluster and attack cluster. Then attack cluster is moved to MLP neural network which classify this records to attack types (DOS (Denial of Service), Probe, U2R (User to Root), R2L. The proposed system has been implemented in windows 7 (64 bit) environment, memory 2 G, processor core i3 and using NetBeans IDE 8.0.2 platform. After testing our IDS using data set, the results of implementation have shown that the detection rate is 99.98%, accuracy is 99.95% and false alarm is 0.08. Simulation results have shown that our IDS provide high detection rate with low false alarms and this is the purpose of our intrusion detection system. For future work this system can be improved to provide higher detection rate for R2l and U2R attacks as our system provide low detection rate for R2L attacks, another improvement to our work is to use an effective feature selection algorithm to select the most important features instead of using all features.

References

[1]. *I.F.Akyildiz, W.Su, Y.Sankarasubramaniam, and E.ayirci. Wireless sensor networks: A survey. Computer Networks, 38(4):393-422, 2002.*

[2]. *M.Burgess. probabilistic anomaly detection in distributed computer networks. Science of computer programing, 60(1): 1-26, 2006.*

[3]. *"Security in Wireless Sensor Networks," A. Perrig, J. Stankovic, and D. Wagner, Communications of the ACM, vol. 47, no. 6, June 2005.*

[4]. *"LAD: Localization Anomaly Detection for Wireless Sensor Networks," W. Du, L. Fang, and P. Ning, IPDPS, 2005.*

[5]. *"Sensor-Based Intrusion Detection for Intra-Domain Distance-Vector Routing," V. Mittal and G. Vigne, ACM CSS, Washington, DC, USA, November 2002.*

[6]. *"Intrusion Tolerance and Anti-Traffic Analysis Strategies for Wireless Sensor Networks," J. Deng, R. Han, S. Mishra, IEEE International Conference on Dependable Systems and Networks (DSN), 2004, pp. 594-603.*

[7]. *"Localized Fault-Tolerant Event Boundary Detection in Sensor Networks," M. Ding, D. Chen, K. Xing, and X. Cheng, IEEE INFOCOM, 2005.*

[8]. *A. P. R. Da Silva, M. H. T. Martins, B. P. S. Rocha, A. A. F. Loureiro, L. B. Ruiz, and H. C. Wong, "Decentralized intrusion detection*

in wireless sensor networks," in Proceedings of the 1st ACM international workshop on Quality of service & security in wireless and mobile networks, Q2SWinet '05, (New York, NY, USA), pp. 16-23, ACM, 2005.

[9]. I. Onat and A. Miri, "An intrusion detection system for wireless sensor networks," in Proceeding of the IEEE International Conference on Wireless and Mobile Computing, Networking and Communications, vol., 3, pp. 253-259, 2005.

[10]. Nikhil R. Pal,Kuhu Pal, James M. Keller, and James C. Bezdek, "A Possibilistic Fuzzy c-Means Clustering Algorithm, " IEEE Trans. on Fuzzy Systems, vol. 13, no. 4, pp. 517-530, Aug. 2005.

[11]. Alaa Eissa, Hazem M. El Bakry, Mamoun H. Mamoun, and Samir Zied, "Intrusion Detection In Wireless Sensor Networks: A survey, " International Journal of Information Science and Intelligent System, vol. 3, No. 4, October 2014, pp. *87-111.*

[12]. Hazem M. El-Bakry, and Nikos Mastorakis, "A Real-Time Intrusion Detection Algorithm for Network Security," WSEAS Transactions on Communications, Issue 12, vol. 7, December 2008, pp. 1222-1234.

[13]. Hazem M. El-Bakry, Alaa M. Riad, Mervat M. Fahmy, and Nikos Mastorakis "Fast Intrusion Detection by using High Speed Focused Time Delay Neural Networks, " Proc. of WSEAS International Conference on Communication and Information, Athens, Greece, December 29-31, 2009, pp.278-295.

[14]. Hazem M. El-Bakry, and Nikos Mastorakis, " A Real-Time Intrusion Detection Algorithm for Network Security, " 8st WSEAS International Conference on Applied Informatics and Communications (AIC '08), Rhodes, Greece, August 20-22, 2008, pp. 533-545.