Data Mining Approach For IDS in WSN

Saroj[1], Imtiyaaz ahmmad[2], Ms. Urvashi3

[1]P.G. Student, Department of computer science & Engineering, I.I.E.T, Samani, KKR, Haryana, India[1]
[2]Asst. Professor, Department of computer science & Engineering, I.I.E.T, Samani, KKR, Haryana, India[2]
[3]Asst. Professor, Department of computer science & Engineering, I.I.E.T, Samani, KKR, Haryana, India[3]

ABSTRACT

Abstract— *This paper proposed a mechanism of intrusion detection created in CWSN. According to varied capabilities and probabilities of suffer attack among sink, CH, and SN, three individual IDSs are designed. An IHIDS for the sink, a HIDS for CH, and a misuse IDS for SN are proposed. Feedback mechanism is used between the sink and CH; HIDS will be retrained for the new type of attacks, which have been detected and classified by IHIDS. For monitoring the status of packets in a Cluster Based Wireless Sensor Network, it is necessary for the packets to establish normal patterns of behavior. Therefore, in this thesis, the rule-based analysis system is used to construct anomaly detection modules and the corresponding rules are defined by experts. Three individual IDSs for the sink, Cluster Head and Sensor Node are planned according to varied capabilities and probabilities attacks that they suffer from. For the sink, an IHIDS is proposed which has the learning ability; it not only decreases the risk of attack, but also learns and adds new classes by learning mechanism in real time when the sink suffers unfamiliar attacks. For Cluster Heads, a HIDS is proposed which has the same detection models as IHIDS, but there is no learning ability in HIDS. Its goals are to detect attacks competently and avoid resource wasting. However, HIDS updates the classes of attacks using the feedback mechanism between Cluster Head and the sink. For Source Nodes, a misappropriation IDS is proposed. A simple and fast method for SN is designed, to avoid SN overwork, and to save resources for the purpose of safety.*

Keywords*: Artificial Neural Network, Integrated Intrusion Detection, KDD Cup, Learning Mechanism*

---

1. INTRODUCTION

The Neural Network Intrusion Detection anomaly intrusion detection system is based on identifying a genuine user based on the distribution of commands she or he executes. This is reasonable because different users tend to exhibit different behavior, depending on their needs of the system. Some use the system to send and receive e-mail only, and do not require services such as programming and compilation. Some engage in all kinds of activities including editing, programming, e-mail, Web browsing, and so on. However, even two users that do the same thing may not use the same application program. The frequency with which a command is used varies from user to user. The set of commands used and their frequency, therefore, constitutes a 'print' of the user, reflecting the task performed and the choice of application programs, and it should be possible to identify the user based on this information. It should be noted that this approach works even if some users have aliases set up as short hands for long commands they use frequently, because the audit log records the actual commands executed by the system. Users' privacy is not violated, since the arguments to a command do not need to be recorded. That is, we may know that a user sends e-mail five times a day, but we do not need to know to whom the mail is addressed. Building NNID for a particular computer system consists of the following three phases:

PHASE 1: **Collecting training data**: Obtain the audit logs for each user for a period of several days. For each day and user, form a vector that represents how often the user executed each command.
PHASE 2: **Training**: Train the neural network to identify the user based on these command distribution vectors.
PHASE 3: **Performance**: Let the network identify the user for each new command distribution vector. If the network's suggestion is different from the actual user, or if the network does not have a clear suggestion, signal an anomaly.

PROBLEM DOMAIN:

A Wireless Sensor Network consists of numerous low-cost, small devices. Usually, when they are deployed to an open and unprotected area, they are susceptible to various types of attacks. In this paper, a mechanism of Intrusion Detection System (IDS) created in a Cluster-based Wireless Sensor Network is proposed. The proposed IDS are an Integrated Intrusion Detection System (IIDS). It can provide the system to counterattack intrusions, and process in real-time by analyzing the attacks. The IIDS includes three individual IDSs: Intelligent Hybrid Intrusion Detection System (IHIDS), Hybrid Intrusion Detection System(HIDS) and misuse Intrusion Detection System. These are designed for the sink, cluster head and sensor node according to different capabilities and the prospects of attacks these suffer from. The proposed IIDS consists of an anomaly and a misuse detection module. The objective is to raise the detection rate and lower the false positive rate through misuse detection and anomaly detection.

2. PROPOSED WORK:

This paper proposed a mechanism of intrusion detection created in CWSN. According to varied capabilities and probabilities of suffer attack among sink, CH, and SN, three individual IDS share designed. The research architecture of our study is shown in figure 1, an IHIDS for the sink, a HIDS for CH, and a misuse IDS for SN are proposed. Feedback mechanism is used between the sink and CH; HIDS will be retrained for the new type of attacks, which have been detected and classified by IHIDS.

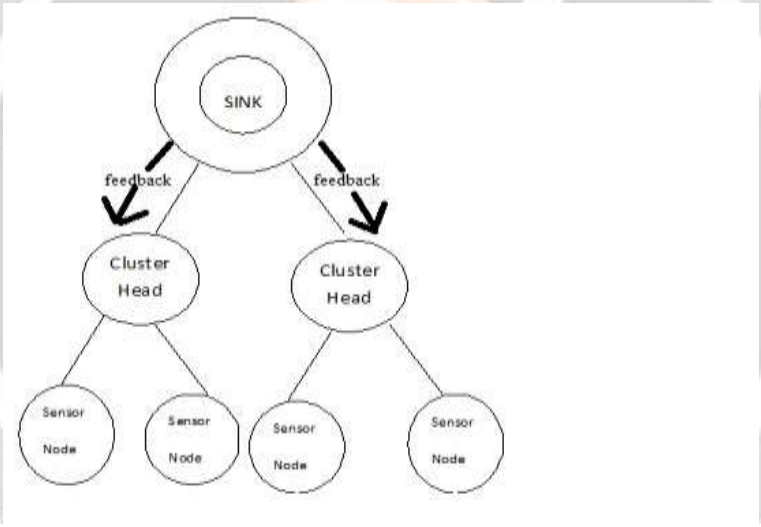


Figure 1: Proposed Architecture

The Figure 2 depicts the system architecture for Intrusion detection. It consists of following components

- KDD Cup'99 Dataset Reading
- Pre-processing & Normalization
- Feature Extraction
- Optimal Feature Selection
- ANN
- Performance Evaluation

Since the single clustering algorithm is difficult to get the great effective detection, "the clustering ensemble is introduced by varying the value of k for the effective identification of attacks to achieve high accuracy and detection rate as well as low false alarm rate". The proposed algorithm produces result of Intrusion detection with high accuracy, no or little false alarm rate and high detection rate

Wireless sensor networks have wide range of applications area such as military applications, field surveillance, Automobiles and many more. Wireless sensor networks consist of various densely deployed sensor nodes inside the application area. Advanced micro-electro-mechanical-systems (MEMS) provides low cost small sized and powerful sensor nodes that are capable of data sensing, data processing and wireless communication and have a limited power battery. Sensor nodes work together to complete the task in time and to provide information accurately. Sensor nodes sense the external environment or application area and send the data to base station located inside or outside the network via single hop or multi-hop. Users access the collected data through some remote access. Sensor nodes work with some limited resources like battery power, memory and bandwidth etc. Wireless sensor networks lifetime depends upon battery power of nodes as every node operation consumes energy, hence node goes out of energy. And it is not possible to recharge or replace the battery of nodes. Therefore, an efficient energy consumption by the nodes is the prime design issue for wireless sensor network from the circuitry of sensor nodes to application level to network protocols. [2]

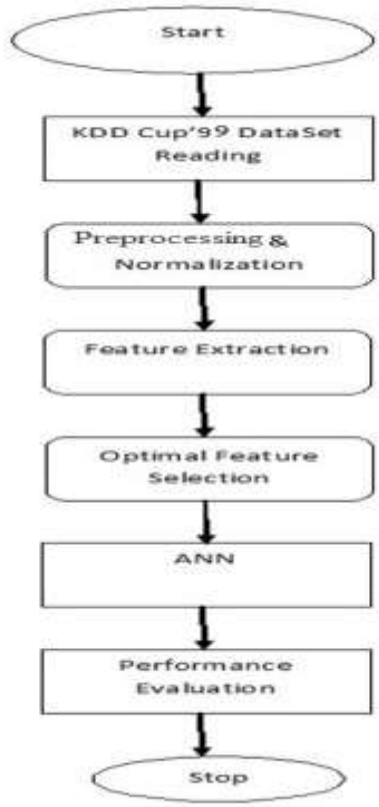


Figure 2: Structure of proposed work

3. MISUSE DETECTION MODEL:

A three-layer BPN is adopted for misuse detection module of IHIDS that comprise of an input layer, a hidden layer and an output layer. Figure 3 shows the structure of the misuse detection model. This use unusual packets, which were determined by anomaly detection module, as the input vector.The number of dealing out units in input layer is determined by the selected features for the packets also, the number of dealing out units in hidden layer is designed through averaging the input layer units and the output layer units. After analysis, there are eight common attacks in CWSN, including Spoofed, Select Forward, Sinkhole, Sybil Attack, Wormholes, Denial of Service, Hello Floods and Acknowledgment Spoofing. Nine dealing out units in the output layer correspond to eight different attacks and one normal behavior, to decide whether the inputted packet is an intrusion and make a categorization.

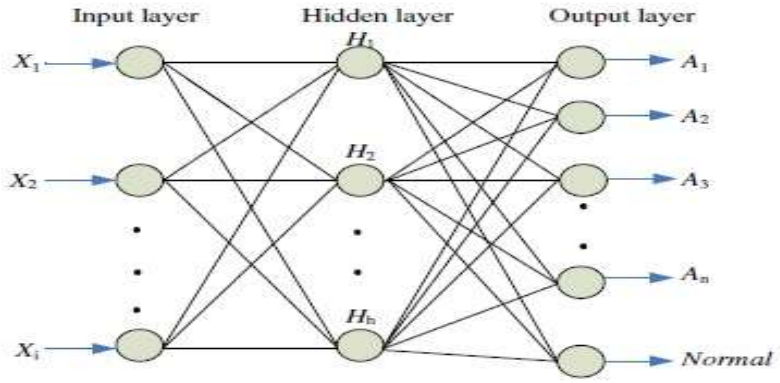


Figure 3: the structure for misuse detection model

Decision making model: The decision making module is used to integrate anomaly detection and misuse detection modules to detect whether the attack is of intrusion or invasion type. The rule-based method is adopted to set up the decision making model, using the set of laws to combine the outputs of two detection models, and its main advantages are that it is very simple and fast.

Intrusion detection for CH: After the above two detection modules are integrated by the decision making module to determine whether there is intrusion and the type of intrusion, and return to the supervisor for follow-uptreatment. The feedback mechanism feeds the data of new attacks, which the learning mechanism of IHIDS provides to the misuse detection model of HIDS for retraining. This way not only does IHIDS get the same performance that HIDS spends additional resources to learn new attacks, but also it saves on resources. When HIDS gets the feedback message from the learning mechanism of IHIDS, the misuse detection model of HIDS is retrained using the data of new attacks at the next training for adding new detection classes.

Intrusion detection for SN: In this paper, a misuse IDS is designed in SN and it is composed of only one model. Misuse IDS determines whether a packet is an attack or not and identifies the category of attack by rules. To avoid SN overburden, and to save resources for the purpose of safety, the intrusion detection in SN is performed through these simple rules. To protect itself from overburden, SN has to use its resources efficiently while detecting attacks and avoid wasting its resources. Because misuse detection has higher accuracy than anomaly detection, the misuse detection is adopted to do intrusion detection.

4. RESULTS AND ANALYSIS:

The implementation results for various parameters are performed as explained below.

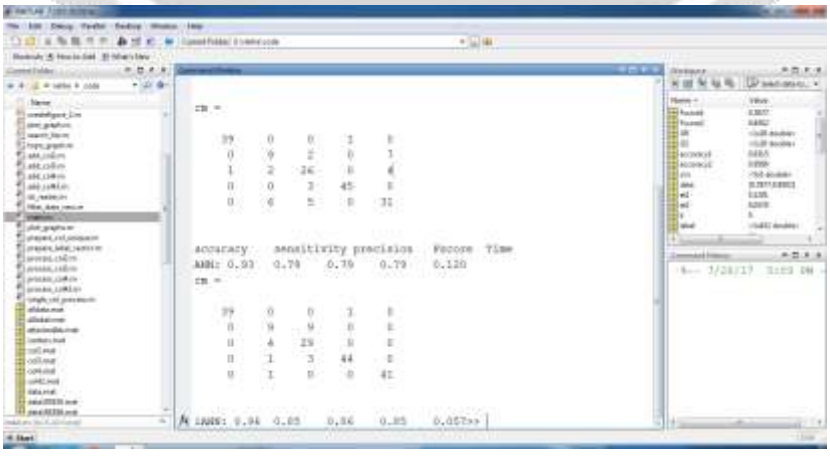
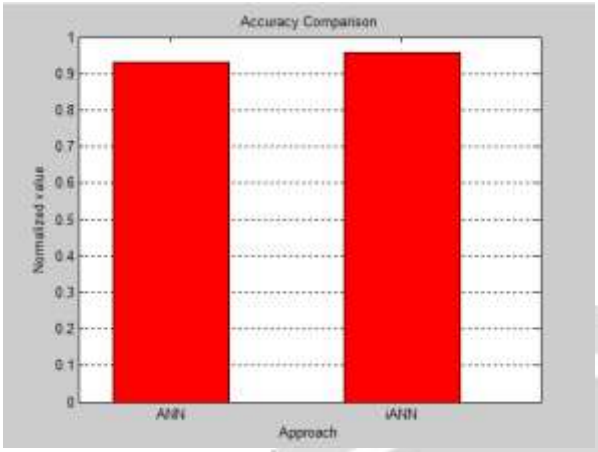


Figure 4: Values of Accuracy, Sensitivity, Precision, F-Score and Time using ANN.

Accuracy: Accuracy is determined by how close a measurement comes to an existing value that has been measured.
Accuracy(% error) = (accepted values- Experimental Values)/Accepted * 100



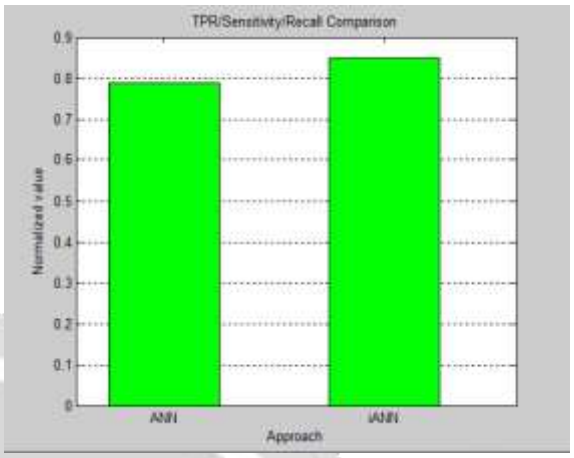
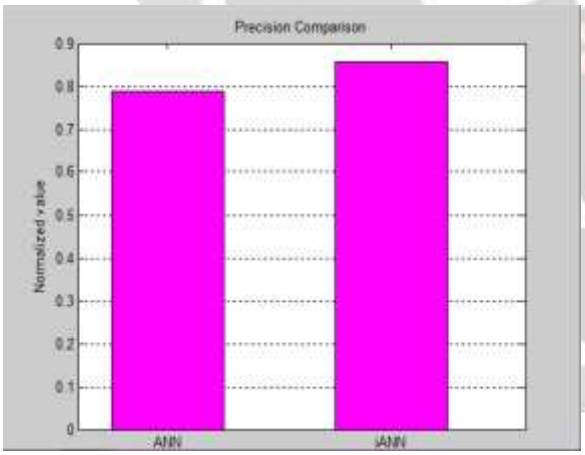


Fig.5:Accuracy Comparison between ANN &iANN ***Fig.6: Normalized value in different approaches used.***

Sensitvity: is the study of how uncertainty in the output of a model can be attributed to different sources of uncertainty in the model input.

Recall/Sensitivity **= (relevant values∩ retrieved values)/relevant values**

Precision: is how close a measurement comes to another measurement. Precision is determined by a statistical method called a standard deviation.
Precision = (relevant values∩ retrieved values)/retrieved values



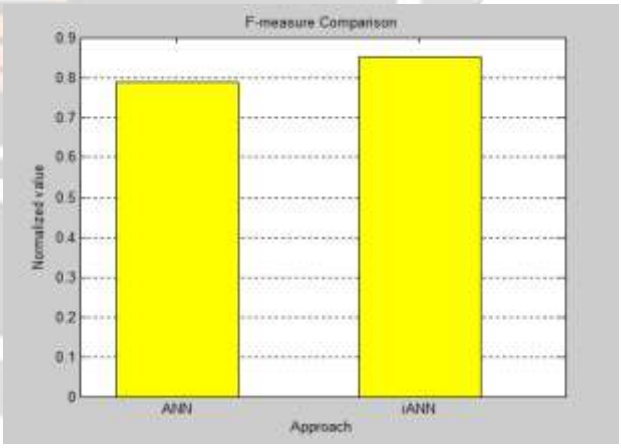


Fig. 7: Precision comparison between ANN & i-ANN ***Fig. 8:F-measure comparison between ANN & i-ANN.***

F-Measure: it is a measure of a test's accuracy and is defined as the weighted harmonic mean of the precision and recall of the test.
F-Measure = 2. (Precision. Recall) / (Precision + Recall)

CONCLUSION:

In this paper, an Integrated Intrusion Detection System is discussed in a varied CWSN. Three individual IDSs for the sink, Cluster Head and Sensor Node are planned according to varied capabilities and probabilities attacks that

they suffer from. For the sink, an IHIDS is proposed which has the learning ability; it not only decreases the risk of attack, but also learns and adds new classes by learning mechanism in real time when the sink suffers unfamiliar attacks. For Cluster Heads, a HIDS is proposed which has the same detection models as IHIDS, but there is no learning ability in HIDS. Its goals are to detect attacks competently and avoid resource wasting. However, HIDS updates the classes of attacks using the feedback mechanism between Cluster Head and the sink. For Source Nodes, a misappropriation IDS is proposed. A simple and fast method for SN is designed, to avoid SN overwork, and to save resources for the purpose of safety. In this paper, the performance of the misuse detection model is evaluated first, which is implemented by BPN. Therefore, to keep the performance for our IDS efficient with a higher accuracy rate, the unknown attacks are classified by the learning mechanism

REFERENCES:

[1] MATLAB Primer, Math Works Inc, 2014.
[2] MATLAB Applications for the Practical Engineer by Kelly Bennett, In Tech, 2014
[3] S. Revathi, Dr.T.Nalini Adam Prugel-Bennett, Gary Wills "Performance Comparison of Various Clustering Algorithm",IEEE, 2013.
[4] Shengyi Pan, Thaomasmarris, "Developing a Hybrid Intrusion Detection System Using Data Mining for Power Syetem", IEEE 2015.
[5] Anna Little, Xenia Mountrouidou, "Spectral Clustering Technique for Classifying Network Attacks", IEEE 2015.
[6] Nur Haryani Zakaria, "Roving K-Means Clustering Using Discretization Technique In Network Intrusion Detection System", IEEE 2016.
[7] S.V. Shirbhate, S. S. Sherkar, V. M. Thakare, "Performance Evaluation of PCA Filter In Clustered Based Intrusion Detection System", 2014 International Conference on Electronic Systems, Signal Processing and Computing Technologiessan, 2014.
[8] Hatim Mohammad Tahir, AbasMd Said, Nor Hayani Osman, Nur Haryani Zakaria, "Improving K-Means Clustering Using Discretization Technique In Network Intrusion Detection System", 2016 3rd International Conference On Computer And Information Sciences (ICCOINS), ©2016 IEEE
[9] N. Deb, M. Chakraborty, and N. Chaki, "A state-of-the-art survey on IDS for mobile ad-hoc networks and wireless mesh networks," IEEE 2011.
[10] J.V. Mulert, I. Welch and K. G. W. Seah, "Review: Security threats and solutions in MANETs: A case study using AODV and SAODV,"IEEE 2012.