ISSN No. 0976-5697

Volume 3, No. 1, Jan-Feb 2012

International Journal of Advanced Research in Computer Science

RESEARCH PAPER

Available Online at www.ijarcs.info

# Artificial Neural Networks in Wireless Intrusion Detection System (ANNWIDS)

M. N. Faruk
Research scholar,
Bharath University, Chennai
justdoit.fff@gmail.com

***Abstract*:** Intrusions are the familiar attacks in current security trend. There are several intrusion detection system enrolled so far, particularly this Intrusion Detection System (IDS) collects the data (activities) at different levels and group them accordingly. It investigates the behavior and nature of commands being executed and its severity level. It carries several layers for processing. The data to be collected and grouped as trails at first layer. The collected data to be configured at the second layer for next level of processing. The Heuristics system and Expert decision making system (EDMS) are the core processes of this approach. Heuristics system does the educated guess of incoming activity by comparing with knowledge Base (KB). Possibly it produces maximum accuracy. Depending upon this output the Expert decision making system (EDMS) does the further level of processing and take final decision. The final decisions are to be concluded based on the three basic parameters (Assurance, Ambiguity, and Divergence).This paper examines the vulnerabilities of wireless network by categorizing the intrusion.

***Keywords:*** Intrusion, Suspicious, Neural network, Heuristics analysis, Expert system

## I. INTRODUCTION

This journal focuses on the discovery of Intrusion detection systems (IDS) that monitors the network or system activities for malicious access or policy violation. It produces the report to security officer and also shows the most often viewed anomalies first.

Today security infrastructure is quite complex and even extreme in nature; generally it includes firewalls, identifications and authentications systems, Access control products, Virtual private networks, Encryption products, virus scanners and more. The firewall includes all of these features, However if any of these feature compromised that will cause serious damage to the system. Out of which IDS are dedicated assistant used to monitor the rest of the security infrastructure.

Security threats in wireless Networks [1,2]

a. Open access to 802.11 networks
b. Unauthorized ("rough") access points
c. Unauthorized use of service
d. Denial of service
e. Mac spoofing and session hijacking
f. Eavesdropping and traffic analysis

There are several techniques has been enrolled to make IDS to more effective. In among the list Rule based expert system for IDS is playing vital role. It act as a (percolate, Trickle, dribble) for incoming data and filter all suspicious events with rule based expert knowledge, it will identify the events (suspicious) depend upon the threshold limit.

The main flaw in rule based expert system is, if the sensitivity level of expert system is high that will cause high false alarm and after possible identification of suspicious events it will forwarded to the security officer or delete by itself.

## II. DEFINITION

Intrusion detection system (IDS) is a security system that monitors the computer and network traffic and analyzes that traffic for possible hostile attacks originating from outside the organization and also for system misuse or attack originating from inside the organization.[1]

An intrusion is a deliberate, unauthorized attempt to access or manipulate information or system and render them unreliable or unusable. To be more precise when suspicious activity is from your internal network it can also be classified as misuse.[1]

In the process of IDS there are two major contradictions may have a chances to occur, false negative and false positive. False negative are riskier than false positive since if there wasn't an attack and IDS classified as suspicious activity is referred as false positive it is not much harm but if there was an attack and IDS doesn't detect it referred as false negative then it can be very disastrous.

The main function of IDS includes 1. Monitoring and analyzing both user and system activities.2.Analyzing system configuration and vulnerabilities 3. Assessing system and file integrity 4. Ability to recognize patterns typical of attacks 5. Analysis of abnormal activity patterns 6. Tracking user policy breach.[2],[3]

Internet security has become more serious issue for anyone connected to the net. Currently the intension of hacking is mainly focusing on E-Commerce sites. The frequency of attacks probes, intrusion attempts is inversely proportional to the difficulty level required to perform such attacks.

It is very difficult to build such IDS system that delivers cent percentage performance and also very difficult robust real – time IDS systems. There are many artificial techniques has been enhanced to reduce the human effort required to build these system.

The IDS can be categorized depend upon the type of attack being performed with the host system. The main aspect of IDS is to alert the security officer IDS can be divided into three broad categories.

### *A. Anomaly Detection Model vs. Misuse Detection Model:*

Anomaly detection model uses recognition techniques for operation sequence.[1],[4] They look for deviation from normal behavior. However they are capable for recognizing

novel attacks. Anomaly detection is carried out by application of results of various scientific methods. There are many methods are possibly implemented like Clustering analyzes, Artificial intelligence methods, scientific mathematical abstraction detection model tech etc.

Misuse intrusion detection model uses signature or rule based detection by which it is very immune and capable of identifying new attacks. Here IDS analyzes and groups all the available information and compares with rule list. Rule list are attack signature given (defined) by IDS. If any new attack has identified then it will update on the rule list. The main drawback is if the security of rule list is compromised that will cause serious damage to the system.

### *B. Network Based System vs. Host Based System:*

This two types of the system are classified depend upon the scope of the security level.[1],[8]. Network intrusion detection system (NIDS) is an independent platform that identifies the intrusion by examining network traffic and monitors multiple hosts. It directly connected to hub or switch and configured for post mirroring or network tab. It has the sensors and placed at inbound and outbound access points, which could be network borders. Sensors will captures all the network traffic both inbound and outbound and analyze the content of individual packets for malicious traffic.

Host based intrusion detection system (HIDS) is setup to detect illegal activities with in the host. The host that identifies the intrusion by analyzing application logs, file system modification, password files, database log list, Access control list. Generally HIDS is a software agent that will audit the logs.

### *C. Passive System vs. Reactive System:*

In the passive system IDS detect al the security violations and analyzes all the suspicious activities, create a log file and alert the security officer.[1],[4]. The reactive system create a log file capture the user activities disconnect the user from network traffic even log off the user from the system.

### *D. Nutshell Analysis of IDS System.[2]*

a. Signature based (Pattern matching)
b. Statistical based.
c. Integrity Checker
d. •Anomaly Detection/Behavior Based
e. Flow Based

### *E. Basic Types of Response of IDS System:*

a. Alteration to the environment
b. Striking back (not recommended)
c. Real time notification
d. Throttling
e. Session Sniping

## III. ARTIFICIAL NEURAL NETWORKS

Neural network is an idea originated in year of 1940, due to practical implications it failed in implementation. [3][9][10]. It got so many barriers that cause major flaw in this field. During the years of 1960-1980, these practical implications are enormously minimized. At last this neural network got maturity and made available in all the purposes.

A neural network contains set of simple units called neurons.[10] Neurons are unique sequence of androids, which recognizes a weighted sum of several inputs according to set of weights. Then it computes Heavisible function to obtain output value. The mass of the output value depends on the number of neurons activated. All the neurons specified by binary values. Neural network got layered architecture; first layer is Input layer which receives the formatted input for further processing. The middle layer is a processing layer which almost receives the filtered input. The algorithm for identifying suspicious events are modeled here and it will also perform an expert system analysis. The output layer delivers the result to the next level processing or to the security officer. The neural network collects the input from various sources and allots one neurons for each. Depending upon the threshold limit each neuron will be activated. [8]

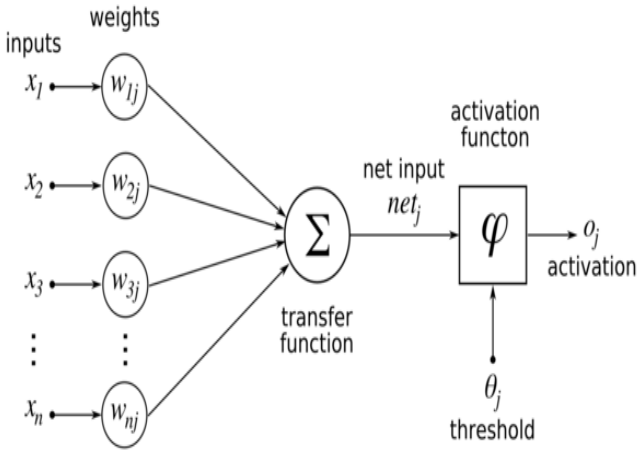



Figure 1.1 The basic function of neural network.

The major difficulties in neural networks are examining the neurons and obtain the desired result. Since each time neurons shows possibly slight difference is resultant weight values. The novel idea is training the networks with samples. The accuracy of the identification is depends on the depth knowledge of a learning algorithm.

## IV. APPROACH – 1

This is the traditional approach of artificial neural networks, here the incoming data will be fed into the filtration process where the data analyzed initial level and filtered with the help of neural network algorithm.[1] After the filtration the formatted output will feed into the expert system analysis process again it will analyze those events for suspicious attacks. The two types of analysis process typically avoid the false alarming.

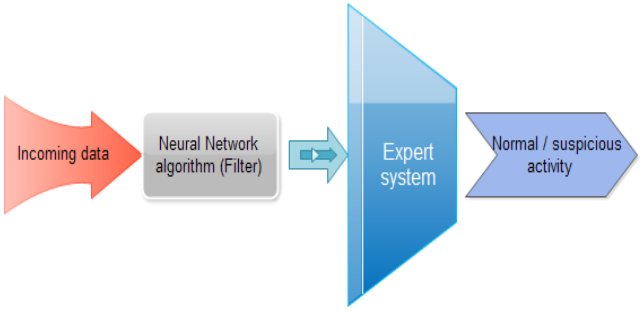



Figure 1.2. The fundamental approach of artificial neural network in Intrusion detection system (IDS)

## V. APPROACH – 2

In the first approach the identification time will be increase if too many similar activities performed at the same time and it will be tedious,[1][3].the simplest form is implemented in approach two, here the incoming data directly fed into the expert system and it configured in such a way that it do filtration as well as identification, the enhancement of this model the slight differ, the expert system feed the final result again to the security officer, that will increase the accuracy of identification. So this security officer system acting as intrusion response system.

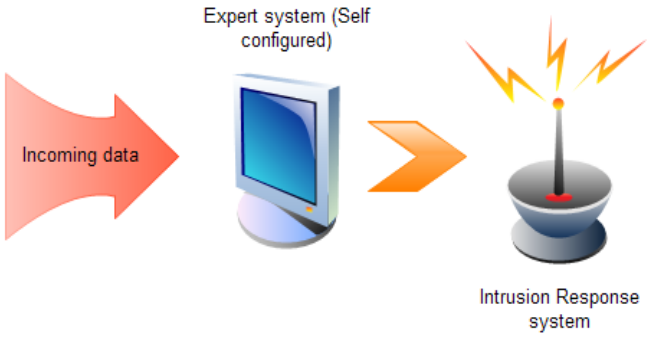



Figure 1.3. The Intrusion response system using artificial neural networks

## VI. HYPOTHESES – INITIAL LEVEL AUDITING

All the user activities to be combined in the form of system logs. The system logs are arranged for next level of auditing process. Auditing records gathered by implementing various auditing mechanisms. Auditing: Conduct systematic reviews to examine carefully for accuracy with the intent of verification.[3][5]

These log files which contain all the sequential process and also all the multitasking events of the operating system. The fundamental perception of networking is *session.* Example login session, it's a semi-permanent interactive information interchange. Once the user login into the session all the sequential commands to be logged at the session file. However session not be created at only once, every time new conversation start and again new session will be created. All the session information temporarily stored in cookie files. The audit records are a dynamic process and mainly focus on multivariate time series.

The audit record contains different set of fields depends on [3] audit imposed to the user. Each field contains unique activity performed by the user. The fields in the audit record have two natures of variables. First finite value for an instance example total no of terminal is in active state at particular period of time. Second contiguous values given by the user for example no of inputs/outputs actions performed by the user.

In order to capture these two types of variables, there are some levels of observations capitalized. The range is declared in numerical form, which is low level to high level symbolic representation.

*a.* ***Keyboard level:*** Monitoring keystroke that will help you to reveal any secret activities that might take place on the computer. There are several techniques has been evolved for tracing (logging) the keys.[3][4] Some of them are hyper visor based key logger it is also called malware based key logger which is placed underneath to the operating system. Other types like kernel based and API based etc, the biggest advantage is it carries very valuable low level data that perfectly suit for neural networks.

*b.* ***Command level:*** In this level it capture all the commands submitted by the user. Generally all the commands are directly associated with system log subroutine. As soon as command executed two special variables added along with the command into the system log.[3][4] First message variable that describes the description of the command can be specified on the command line. Second file variable read each line of the command is being executed. By implementation these two variables got low level information and much suitable for neural networks.

*c.* ***Session level:*** As soon as session getting ends up the monitor calibrate the entire log file, it holds information about length of the session, over all CPU memory usage and time of log in etc, however it is not good for real time usage since hackers steal all the information during the session itself.[3]

## VII. HYPOTHESES- INITIAL LEVEL PROPOSITION

The task performed by the computer is only accomplished by the commands given by the user;[5] this is a fundamental approach of computer operations. In contrast the way of user commands to the computer is not dramatically change in nature. By habits computer will receive similar set of commands and keep using it as its habitual work actions.

The user activity which is performed on the computer are arranged in a sequence, the sequence tells us the user behavior, the perplexity is when the user changing the activity sequence, it is very difficult to observe the user behavior, this diffusion state is referred as stochastic process. Observing the user behavior is complex task since user activities may not same all the time, it will vary time to time.

Assume that user logging in every time, Generally the first and foremost activity is checking mails ,reading news papers and playing music etc, these activities are stored in a log file by having probability of all the activities performed, the sequential order list will be generated. Another idea of generating the audit record samples. Audit record gathers several data samples that qualify the given event. The sample which includes all the data of qualifying attributes. However sometimes it is quasi-stationary. The data qualifying attributes will increase more accuracy in samples. Example computer memory usage, file system, input-output load etc.

## VIII. MODEL - IMPLEMENTATION

So far we discussed about the idea and some approaches to make perfect IDS. In order to make an expert system novel experience is needed. There are some practical approaches implemented here in this model.

It is quite difficult to evolve neural network technique directly into IDS. Before the data feed into neural analysis it has to be forecasted and intake possible audit trails, since entire system data cannot be taken into consideration.[1][5]

In practical there are some five modules has been implemented in Intrusion detection system (IDS).

***a.*** ***Data Collection:*** By collecting different type of audit trails and arranging them in sequence then feed into the next level of processing. This module simplifies the grouping and arrangements of audit trials.

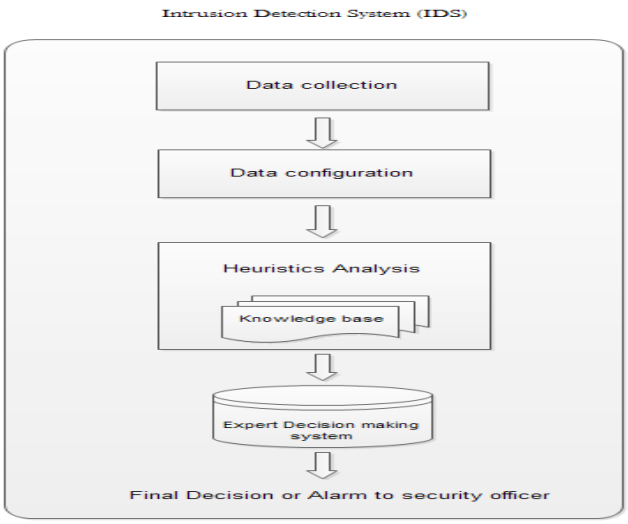



Figure 1.4 the layered architecture for Intrusion detection system using artificial neural networks.

***b.*** ***Data Configuration:*** This module configured the data in common format that suit for further level of heuristics processing. Here the data is converted into binary form and encrypted to provide the possible security.

***c.*** ***Heuristics Analysis:*** Heuristics uses the past experience to make educated guesses about the present. In security point of view heuristics does the following process like file emulation, file analysis and generic signature detection. The incoming data will decrypt and analyzed compared with knowledge base. However it does not give maximum accuracy (it may trigger false alarm).

***d.*** ***Expert Decision Making System:*** This part of the system contains the main process of IDS, it already receives the data with maximum identification and it does the rest of the process. Finally it will decide the whether the current activity is intrusion or normal activity. The output of this module sent to Intrusion prevention system or alarm to the security officer.

## IX. EXPERIMENTAL RESULTS

After successful implementation of this model, neural net obtained all outputs and stores result of each audit data.

## X. PRELIMINARY TEST

We conducted a test with anonymous user who using Linux system and entering sequence of commands. The audit record contain all the command details it also includes system information like memory usage, total bandwidth allotted for each session etc, If the user working on integrated development environment user may work with different types of windows. It is very difficult to capture all the activities performed by the user.

The initial experiments conducted with the user who entering sequential stream of commands. Single neurons will be allotted for each command. For example if a person entering 100 commands 100 neurons will be allotted. In order to get the accuracy of allotment for each neurons probability will be assigned 0 or 1. [6]

The perfect observation of neurons for each command will be examined for next level of processing. The number of activation of each output neurons will be calculated. If the user commands gets maximum will be considered as intrusion.[6][4].

In order to classify output neurons there are primary measures will be taken into consideration.

***a.*** ***Assurance:*** If activation is over 0.5 which means the command is assured. It shows good sign with the user. All the legitimate commands are comes less than 0.5 activations. However it does not give accurate result, since the commands are not in common nature. If any command invokes any secure file, it does not mean as intrusion. In order to identify the intrusion the density of probability of occurrences is calculated. By average the maximum commands is got over 0.5 activation is referred as normal behavior with the user. If this prediction is wrong it will cause serious damage to the system.

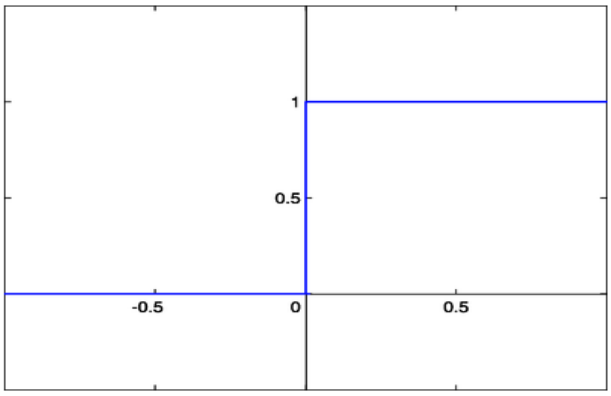



Figure 1.5 Activation threshold limit

***b.*** ***Ambiguity:*** If the sequence of commands activation stays at 0.1 level and the output neurons doesn't show any difference,[5] it means the neural net unable to determine the nature of the command. This is possible only when the configuration is incorrect with neural net.

***c.*** ***Divergence:*** The big fluctuation with maximum activation between 0.1 to 0.5 level, the neural net determines that the command likely to be an intrusion. However it cannot be defined only with help of one fluctuation with two commands. If the probability of occurrence is too high, which likely to be an intrusion.

As new dimension of observation the neural network has to be well trained and identify the activation level of each command. To enhance this model we received around 2500 commands from Linux network user. While training there is lot more opportunity for prediction error. The commands are grouped into four sections. The first row says total number of commands executed and last column describes the prediction error. The prediction error must be very less at the end of the training.

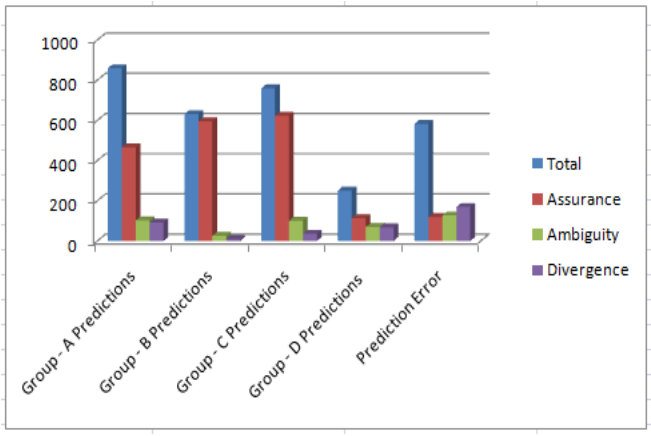
The key observation is to count maximum activation with the given commands. The maximum activation shows the level of confidence with each command. If the total activation increases the gives maximum accuracy. While training the divergence will be high and that will also cause

prediction error. The divergence count supposed to be less for accurate identification.

As we have closer analysis of these results, the prediction error caused because of *noisy* commands. The *noisy* commands are randomly undeterministic sequence. Example date and time commands. It may also possible when similar actions in two different commands.

Table 1.1 prediction of Sequence stream of commands

| Convergence | Group-A Predictions | Group-B Predictions | Group-C Predictions | Group-D Predictions | Prediction Error |
|---|---|---|---|---|---|
| **Total** | 859 | 631 | 759 | 251 | 583 |
| **Assurance** | 456 | 595 | 623 | 113 | 119 |
| **Ambiguity** | 102 | 26 | 100 | 70 | 127 |
| **Divergence** | 91 | 10 | 36 | 68 | 169 |
| **Avg.No. of activation** | 0,243 | 0,163 | 0,099 | 1,078 | 0,020 |

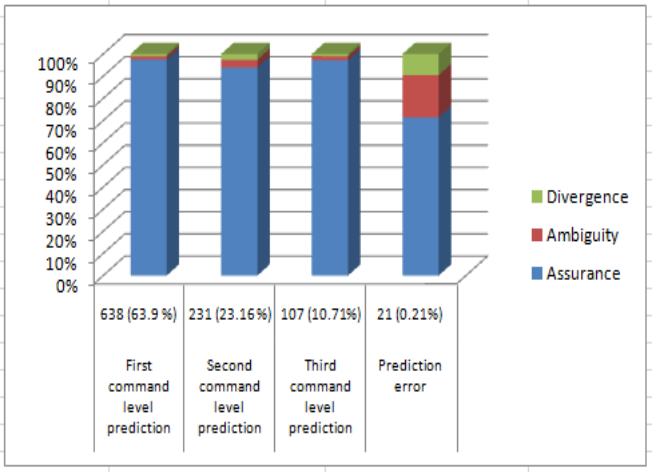



Graph 1.1 Graph notations for group of commands.

This enhancement experimented in network session. Session is a temporary time range of network usage. The network session starts when a user logging in. Initially the single session given to this model for training. While training there are 759 commands taken into consideration for the session. These commands are sectorized and applied to this model. The main goal of this model is to get minimum prediction error. The level confidence only recognized by the number of activation. However confidence level cannot be identified directly.

Table 1.2 Command level prediction for first session.

| Session (1) | First Command Level Prediction | Second Command Level Prediction | Third Command Level Prediction | Prediction Error |
|---|---|---|---|---|
| **Total** | 638 (63.9%) | 231 (23.16%) | 107 (10.71%) | 21 (0.21%) |
| **Assurance** | 622 | 217 | 104 | 15 |
| **Ambiguity** | 10 | 8 | 2 | 4 |
| **Divergence** | 6 | 6 | 1 | 2 |




Graph 1.2 Graph notation for session one

The prediction error is independent to the no of commands entered in the session. That gives stability of this model.

Table 1.3 Command level predictions for all sessions.

| Sessions | No. of Commands | First Command Level Prediction | Second Command Level Prediction | Third Command Level Prediction | Prediction Error |
|---|---|---|---|---|---|
| Session 2 | 1256 | 997 (79.37%) | 83 (6.60%) | 23 (1.83%) | 79 (6.28%) |
| Session 3 | 538 | 403 (74.90%) | 43 (7.99%) | 16 (2.97%) | 48 (8.92%) |
| Session 4 | 610 | 538 (88.19%) | 57 (9.37%) | 19 (3.11%) | 59 (9.67%) |
| Session 5 | 387 | 239 (61.75%) | 23 (5.97%) | 8 (2.09%) | 28 (7.23%) |
| Session 6 | 637 | 517 (81.16%) | 56 (8.79%) | 18 (2.82%) | 57 (8.99%) |
| Session 7 | 987 | 876 (88.75%) | 88 (8.91%) | 8 (0.81%) | 78 (7.98%) |
| Session 8 | 679 | 506 (74.52%) | 56 (8.24%) | 12 (1.76%) | 23 (3.38%) |
| Session 9 | 386 | 287 (74.35%) | 27 (6.99%) | 3 (0.77%) | 18 (4.66%) |
| Session 10 | 549 | 418 (76.13%) | 46 (8.37%) | 4 (0.72%) | 19 (3.66%) |

All the command level inputs are gathered together to conclude with single entry in the audit record. For example if the first command declines from 85% to 77% then the audit record will enter 81%. The prediction error is around 10% which means network is not trained properly and it will rectify in further observations. The prediction error with confidence happens only if equal or less than 3%. The accuracy level is purely depends on the sensitivity level of heuristics system and expert system.

## XI. CONCLUSION

Wireless networking has been an issue of security since its creation. Even though so many technologies invented for wireless security, still intrusion is unbeaten in time to time. This paper investigates a new approach to strengthen the security system for intrusion attacks. The neural network provides the fundamental idea, which helps in classification of commands based on their nature. However neural network alone cannot give perfect diagnosis for intrusion.

The Expert decision making system (EDMS) and heuristics system also playing vital role.

With this system we got inputs from normal network user. In order to improve the security level of these system Artificial intelligence guidelines required in Expert decision making system (EDMS). By implementing the rule based expert system in heuristics analysis that will provide maximum security.

## XII. REFERENCES


[1] Sattar B.Sadkhan, "On Artificial Intelligence Approaches for Network Intrusion detection systems", In MASAUN Journal of computing volume 1, Issues 2, September 2009, page 236-238.

[2] Slobodan Petrovic, "vulnerabilities in wireless networks and intrusion detections", In Article of Telektronikk,2005, page 86-89

[3] Iftikhar ahmed, Azween B Abdullah, Abdullah S, Alghamdi," Artificial neural network approaches to Intrusion detection : a review ", In telecommunication and informatics book as ACM guide included in ISI/SCI web of science and web of knowledge, page 200-204.

[4] Beqiri E., Lee .S.W; Draganova C and Palmer-Brown .D(2010)," A Neural network approach for Intrusion detection system", 5th conference in advances in computing and technology(London, United Kingdom, 27th January) page 209-217.

[5] Herve DEBAR, Monique BECKAR, Didier SIBONI, " A Neural network component for an Intrusion Detection system", In IEEE ASAP Magazine, 1992, page : 240-244, 245-248.

[6] Timothy James Stich, Dr Julie K Spoerre & Dr.Tomas Velasw," The Application of Artificial neural networks to monitoring and control of an induction hardening process", In Journal of Industrial technology, vol 16, Number 1 , Nov-1999 to Jan 2000. Page 2-5.

[7] Jake Ryan, Meng –Jang Lin, Rioto Mikkalainen, "Intrusion Detection with Neural Networks", In Advances in neural information processing systems, Cambridge, MA, MIT Press, 1998, Page 2-7.

[8] Alan Biven, Chandrika Palagri, Rasitha Smith, Boleslawszymansk, Mark emberchts, " Network based Intrusion Detection using neural Networks", ANNIE – 2002, St. Lousis Vol 12, ASME press New York, NY 2002, page 579-584.

[9] XIN YAO," Evolving Artificial Neural Networks:", In Proceeding of the IEEE volume 87, Number 09, Sep 1999, page 1434-1436.

[10] Lee.K Chang. J , and Chen. M. "PAID: Packet analysis for Anomaly Intrusion Detection". In Advances in knowledge Discovery and Data mining, National Taiwan University, 2008 PP 626-633.


### Short Bio Data for the Author

M.N.Faruk received engineering B.Tech degree from Department of Information Technology, Kamban Engineering college, Anna university in 2006. And M.E degree from Department of Computer science, Arunai Engineering college, Anna university in 2011. Presently he is working in Arunai College of Engineering as a Assistant professor. He is pursuing ph.D at Bharath University, Chennai under the guidance of Dr. M. Rajani. He has published 2 research papers at national and International level. His area of interest is Wireless security, Data mining.