Hybrid Approach of Intrusion Detection based on Sequential Feature Selection, EM Clustering and Decision Stump Classification

Gulshan Ansari
Department of Computer Science and Engineering
ASCT, RGTU,
Bhopal, India

Tanveer Farooqui
Department of Computer Science & Engineering
ASCT, RGTU,
Bhopal, India

ABSTRACT
Other than the emerging IT sector, security is still being a major issue for various companies. Various companies are suffering from several types of threats these days like viruses or intrusions, etc. There are various types of techniques have been applied by the companies like for detection of intrusion and also for providing prevention to the system in order to secure the companies against these types of intrusions. The technology of Intrusion-detection can have a lot of problems, such as low performance, low intelligent level, and more false-negative-rate, high-false-alarm-rate, and so on. The purpose of the IDS is to detect attacks. Objective of this analysis is to describe the phases of the development of concepts of the IDS along with its significance for the researchers and the research-centers, military domains, security and also in order to determine the significance of IDS categories, its classifications, and in which area applied the IDS in order to decrease the threats of network.

Keywords
Feature Selection, EM Clustering, Decision Stump Classification, Intrusion Detection

1. INTRODUCTION
The concept of Intrusions within the computing domains is most commonly unwanted type of malignant function or working which is emerging since the beginning of the computing technologies. And various security policies and approaches have also been developed within the last few years, but with the increase of Technology use, the rate of security threats have also been raising. As the entire world is dependent over the computers, being directly or indirectly, it is a very important issue to prevent the malicious activities and threats that can hamper the computing infrastructures. Considering the emergence and simultaneously analyze the explosion within the networks of computer have enhanced the influence of social-web. And the pattern in which content is distributed and got accessed, is also be the basic of the latest global tradition of working, also it is affecting and merging the parts of the private and business routines of life.

In this describe the concept of detection of intrusion in MANET environment is also a complicated and troubling work basically because of the dynamic behavior of the MANETs, and its highly limited nodes having lack of the centralized monitoring concepts. Traditionally the IDS are not easily get implemented within MANET. Some latest techniques are required to be introduced or may be some available techniques are required to be applied within MANET. In this paper represented some issues of IDS for the MANET area and also analyze some solutions suggested here. A boundlessness of approach for the misuse type of detection also the anomaly type of detection has also been applied. The rule based approach usually does not generate a large number of false alarms since it is based on rules that identify known intrusions but it fails to detect new types of intrusions as their signatures are not known.

According to analysis the susceptibility estimation and the prevention of intrusion or its detection are the aspects of the security-management domain of IT. Though, because of the latest establishment along with regular spreading the connectivity of network for security-management of IT is also suffered some issues yet, also it requires an efficient technique for the proper functioning. However the fast type of networks may have provided the rise to the activities of business over the Internet as they may also have to describe the completely latest range of the computer's threats that are not known earlier.

And these types of threats are also associated with the issues of security and the cyber-crimes. Currently several online type of applications are there those are within the attack of various viruses, different type of malwares, some malignant contents, affected by Trojans, etc. So an efficient intrusion can be tried by the hackers in order to affect any type of organization so that is not only affect the privacy of the information and also it may causes the financial losses whereas the significant damage made is the loss of the integrity and the trust. Users will no more be interested to do the business with those organizations which are not able to secure their information.

In this paper, based on a comprehensive analysis for the current research challenges in intrusion detection, In mobile ad-hoc network all nodes can roam independently which indicates nodes which does not having fixed architecture can be attacked and compromised easily. As it is very difficult to keep track of any mobile node in a huge global network, attacks from such a nodes are difficult to detect and can be more dangerous to network. So in ad-hoc network mobile nodes and infrastructure cannot have trust based relationship. As an important technique in the Defense-in-depth network security framework. Intrusion Detection has become a widelystudied topic in computer networks in recent years[4].

The related issues of the machine learning mechanism are the ways systems get enhance automatically their efficiency along with the raise of practice, which is consistent with that of the IDS [7].

2. SECURITY ISSUES IN MANET

Different characteristics of MANETs make conventional IDSs ineffective and inefficient for this new environment. Consequently, researchers have been working recently on developing new IDSs for MANETs or changing the current IDSs to be applicable to MANETs. There are new issues which should be taken into account when new IDS is being designed for MANETs.

2.1 Routing

Routing protocols are used to find the optimal path from source to destination node. Routing protocols are used to exchange the routing information. These are very important in MANET where topology changes very frequently due to mobility of nodes. Mobility, Bandwidth constraints, Hidden and exposed terminal problems and Resource constraints [7] of the nodes are some of the challenges which need to be addressed while designing routing protocols for MANET.

2.2 Lack of Central Points

Due to the lack of central point of control, MANETs are more vulnerable to routing attacks as compared to other networks. Although security has long been an active research topic in wire line networks, the exclusive features of the MANET have present a latest set of the nontrivial type of issues to the security model.

2.3 Data forwarding security issues

Protecting the network layer in a MANET is an important research topic of wireless security. The core functionalities provided within network-layer are the routing or the packet forwarding, malicious attacks on either of them will disrupt the normal network operations.

2.4 Dynamic topology

In MANET node may join or leave dynamically. As node moves frequently establishing trust among nodes are very difficult.

2.5 Error-prone channel state

The characteristics of the links in a wireless network typically vary, and this calls for an interaction between the routing protocol, if necessary, find alternate routes.[4]

2.6 Mobility

The network topology within the ad-hoc type of wireless networks is highly dynamic due to the movement of nodes; hence an on-going session suffers frequent path breaks. This situation often leads to frequent route changes.

2.7 Hidden terminal problem

This issues is refers to the collision of packets at a receiving node due to the simultaneous transmission of those types of nodes which are not follow the direct types of range of transmission from the sender, but are within the transmission range of the receiver.

2.8 Energy-restricted functions

Few nodes or may be all types of nodes within the MANET are dependent over the batteries or some other expandable ways for providing energy. And to these types of nodes, very significant type of design criteria of system for the optimization is considered to be the conservation of energy.

3. INTRUSION DETECTION SYSTEM

An intrusion can be termed as an unauthorized entry to another's property or area, but in terms of computer science, it is the activities to compromise the basic security objectives of computers viz. confidentiality, integrity, and privacy. The process of Intrusion-Detection is basically monitoring all events that are found within the computer or the network also it analyzes those events for the marks of some possible intrusions or the threats and also any type of violations of the security policies of computer, some acceptable application of policies or may also analyze the standard policies of security. The IDS may be described as the tools, resources, and methods in order to support in identifying, assessing and also it report the illegal or invalid activity of network. The process of Intrusion-detection is basically being the part of the entire mechanism of protection which is installed over the system or the device also it is a combination of different measures of protection.

IDS may support the systems of information to create for and it works with the attacks. It also performed this through gathering the information among various different types of systems and also from various sources of network and also it analyzes the information to resolve some possible issues of security.IDS may also detect the attacks also some anomalies in the network, and thus are becoming very important. IDS are useful in detecting successful intrusion, and within checking out the traffic of network and the attempts to break the security. Intrusion detection is the practice of observing and examining the actions going on in a system in order to identify the attacks and susceptibilities.

Depending on the detection techniques used, IDS may also be categorized in 3 major types [7] as follows

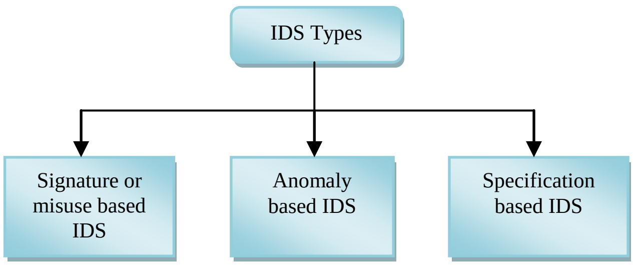

Figure.1: Intrusion Detection Types

3.1 Mis-Use Based IDS

Most commercial IDS look for attack signatures:

Specific patterns of network traffic or activity in log filesthat indicate suspicious behavior are known as knowledge-based or misuse detection IDS. The Signature-oriented system may be referred as the mis-use-detection or may be knowledge-oriented mechanism. The Mis-use systems of detection can compare the latest working of any host system or any network which is checked on the basis of signatures of the known type of attacks. And in case the activities got matched with any of available signatures then alarm is generated.

3.2 Anomaly Oriented mechanism of Detection

This mechanism may try to represent the usual functioning of the system. And any type of event that may break this model's behavior if found is referred as the suspicious activity. This is objected to find out the signatures or marks which never ensure the usual functioning. Those signatures or marks which never considered as the normal functioning are termed as the anomalies. Events in the anomaly oriented engine for detection are created by any type of functioning which may not belong to the pre-defined or the accepted type of model of the behaviors.

3.3 Specification oriented IDS

In specification-based detection [7][8], the correct behaviors of critical objects are manually abstracted and crafted as security specifications, which are compared with the actual behavior of the objects. Intrusions, which usually cause object to behavior in an incorrect manner, can be detected without exact knowledge about them. So far, specification-based detection has been applied to privileged programs, applications, and several network protocols.

IDS may collect and analyze the information among various systems and the sources of network for the signs of the intrusions. And the IDS may also be the host-oriented or the network oriented mechanism. The host-oriented IDS are situated within the servers to examine the internal interfaces and the network-oriented IDS can monitor the network traffics for detecting intrusions. The Network-oriented approach of IDS may performs packet logging, analysis of traffic of the IP network, and tries to discover if an intruder is attempting to break into the network

4. LITERATURE REVIEW

In this paper [11] suggested the behavior-rule-specification-oriented approach of IDS for detection of intrusion within the medical instruments or devices that are embedded within the MCPS. In this demonstrated the functionality along with the VSMs and also represented that probability of detection of those medical-device may reaches the those system which most get attacks whereas restricting the probability of false alarm to be less than 5 percent for the hopeless attackers and also less than 25 percent for the random type and the opportunistic type of attackers within the wider range of the noise level of network. By this analysis represented that the suggested specification dependent approach of IDS may proved to be efficient over available approaches that are dependent over the anomaly type of IDS.

Within this paper [12] describes the most essential domain of research within the concept of the Detection of Intrusion which is associated with the implementation of AI approaches. This suggested mechanism may represent a latest technique of the IDS dependent over the ANN concept. And the MLP infrastructure is also applied in this for the IDS purpose. So the efficiency and the evaluations have been carried by the use of the group of the benchmark-data among the KDD data-set. This suggested system may find out the intrusions and also it classifies those intrusions into groups for analysis.

In this paper [13] suggested after implementation researchers are able to increase the detection rate and decrease the rate of false-alarm of the IDS through merging the two type of data mining algorithms C4.5 Decision Tree and SVM. Comparison is done using various parameters like Detection-rate, Accuracy, false-Alarm-Rate, etc. Creating an efficient IDS models along with good accuracy and real-time performance are essential. However, other kinds of preprocessing techniques and data mining approach such as AI,NN approach may be tested for a better detection rate in the future research in IDS mechanism.

According to this paper [14] suggested the fuzzy genetic algorithm that was used in [7, 8] may have the higher rate of detection as compare to the decision tree algorithm in most cases, and it was good at detecting unknown attacks. It had a higher detection rate than the traditional genetic algorithm that was used in [6]. The genetic algorithm in [6] had a high detection rate for denial of service attacks. When compared with the winning entry of the KDD99 Classifier Learning Contest, it was shown to have a better detection rate for both denial of service and user to root attacks. This paper showed that the use of genetic algorithms and fuzzy genetic algorithms in intrusion detection are effective ways of detecting attacks

Within this paper [15] suggested the real world network oriented IDS approach of PSO-Discretize-HNB along with high accuracy. The proposed type of NIDS may merge the PSO approach of feature-selection approach and IEM discretization along with the HNB classifier. And applied the NSL-KDD data-set for intrusion of network type benchmark was used for conducting several experiments to test the effectiveness of the proposed the NIDS. Also, a comparative study with applying IG feature selection and IEM discretization along with the HNB type of classifier have been obtained. Also in order validate the suggested PSO-Discretize-HNB approach of detection of intrusion of network it is also get compared along with the various feature-selection approaches like the PCA, gain-ratio, etc.

In this paper [16] described a latest technique through applying the approaches of DM like the neuro-fuzzy and the radial oriented SVM in order to support the IDS for obtaining the higher rate of detection. This suggested approach can have the four types of steps like basically the k-means clustering process is applied in order to produce various types of training sub-sets. After that dependent over the achieved training-sub-sets, various neuro-fuzzy schemes have got trained. And after this, the vector for the SVM approach of classification is applied and finally, the classification is done by the use of radial-SVM has been carried out in order to find out the intrusion either occurred or not. And for explaining the relevance and the ability of latest technique, the outcome of the experiments over the KDD-CUP-1999 have been represented. By the results of experiment have presented that the suggested latest technique performs better as compare to the BPNN, multiclass-SVM, and so on.

According to this paper [17] it is analyzed that it is complicated to design an effective type of mechanisms for the detection of intrusion in order to secure the MANET against the attacks. Along with the enhancements of technologies and reduction in the costs of hardware users are observing the latest trend of extending the MANET within the industrial domains. So in order to adjust with those trends, it is strongly considered that this is more effective to refer its issues regarding security. Within this paper, suggested and also apply the latest type of IDS referred as the EAACK which is specially developed for the MANET environment. Also it compared with the existing algorithms, this EAACK approach may represented the higher rate of malignant-nature of detection within some situations whereas is never influence the performances of network.

5. PROPOSED WORK

This section is dealing with the structure of the proposed work. In this section, the proposed work is described in two phases. Firstly, this section shows the architecture of the proposed work. This proposed architecture is shown in figure 1.

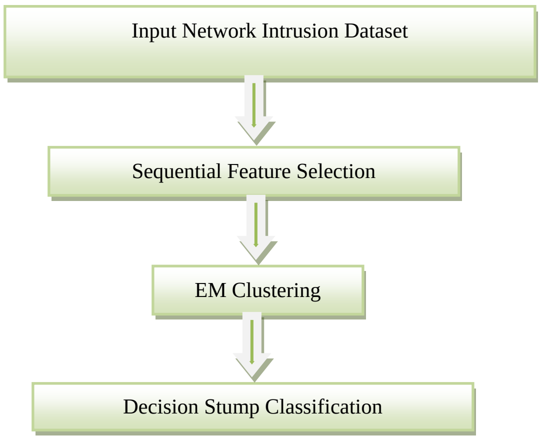

Figure 2: Architecture of the Work

This architecture shows that the intrusion detection is done on the dataset. Firstly, proposed work reads dataset then it uses sequential feature selection method to select the various attributes of the dataset which are significant or can say which could play important role in intrusion detection. In the next the step EM clustering is used to separate the malicious data from the normal or generic data. Finally, the rest of the dataset is classified into various attacks through Decision Stump Classification.

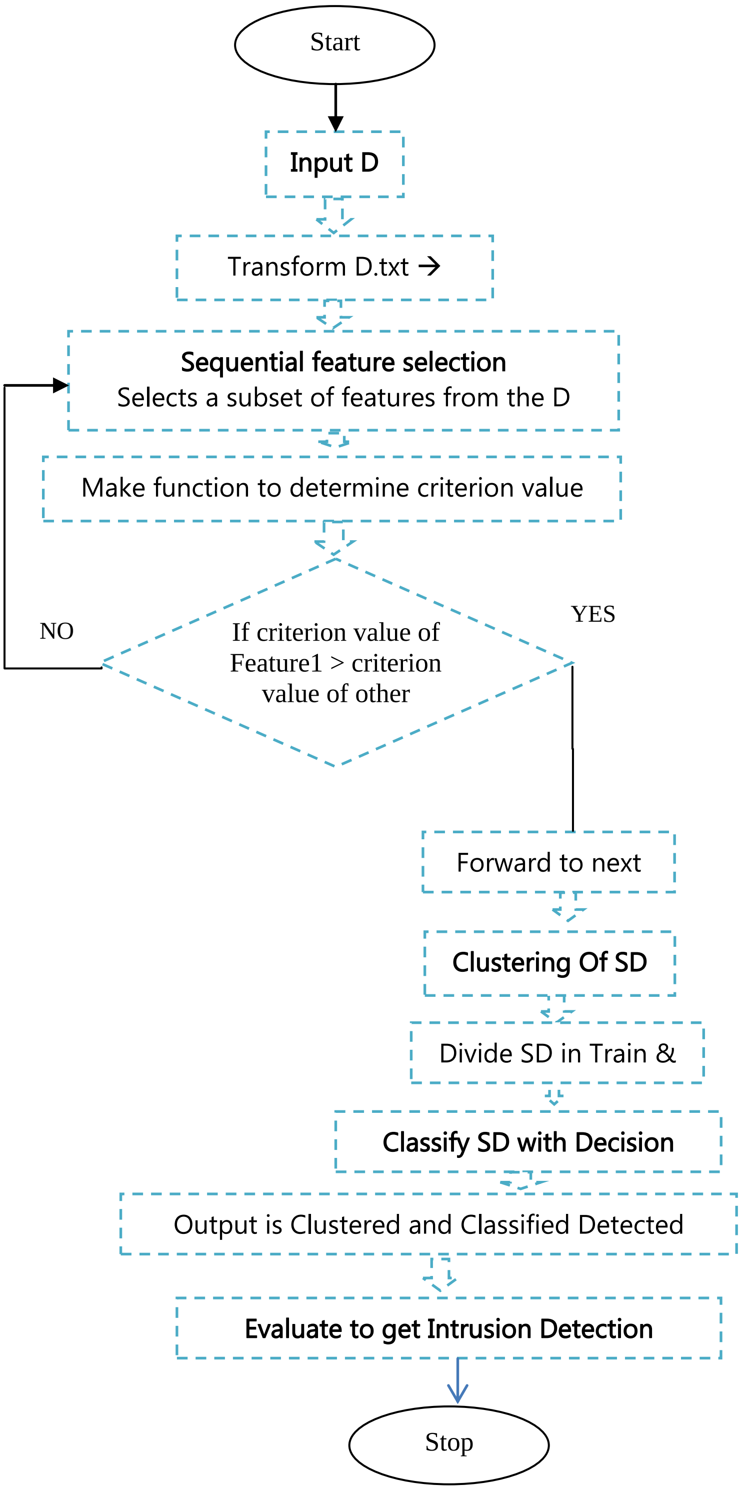

Figure 3: Flow chart of this proposed work.

Figure 2: shows the details working of the proposed work with the help of the flow chart.

1. Start
2. Input D(Network Intrusion Data Set)
3. Convert D text to D mat
4. Sequential feature selection (SFS)
    a. *SFS selects a subset of features from the D matrix X*
    b. *Make function 'criterion = fun(XTRAIN, Ytrain, XTEST, Ytest)'*
    c. *Function determine criterion value for each feature*
    d. *If criterion value of feature1 > criterion value of other features*
    e. *feature1pass forward*
    f. *else feature1 return to point 'a'*
    g. *Function reaches its limit and stop process*
    h. *SFS out logical value, 1 for in and 0 for out*
5. SD(Selected Feature Dataset) formed
6. Expectation Maximization clustering
    a. *Input is a matrix X with m rows and n columns. Element in place (i, j) is denoted by X_{ij} and NI_types (types of intrusion) is k (number of clusters).*
    b. *Define the initial belonging probabilities. The result is a m*t matrix TC and element TC_{it} in place (i,t) is the probability that rows I belongs to cluster t.*
    c. *First all elements of the matrix P are set to be zero and then on each row one element is set to be one to create k clusters of the equal size.*
7. SD observations with its associate cluster parted equally Train and test
8. Test and Train classify by Decision Stump Classifier
    a. *If all output values are the same in SD, return a leaf node that says "predict this unique output"*
    b. *If all input values are the same, return a leaf node that says "predict the majority output"*
    c. *Else find attribute X with highest Info Gain*
    d. *Suppose X has distinct values n_X*
    e. *Create and return a non-leaf node with n_X children*
    f. *The i'th child should be built by calling Build Tree(SD_i ,Output) Where DSi built consists of all those records in DataSet for which X = ith distinct value of X*
9. Output is Clustered and Classified Detected Network Intrusions
10. Evaluate results to get correct performance of Intrusion Detection
11. Stop

Figure 4: Algorithm of proposed work

6. RESULT ANALYSIS

This section is divided into three main parts which will find the result and the system and data structure on which these results are found.

1. System Configuration with Software used

Model:	Sony Vaio
Processor:	Intel® Core™ I5-2450M 2.5GHz
RAM:	4GB
System Type:	64 Bit Operating System
Windows Edition:	Windows 10 Home Basic
Matlab	R2014a
WEKA	Weka 3

2. Data set

The dataset chosen and taken is KYOTO 2006+. This dataset is having 14+1 attributes where 14 attribute are general attributes and last attribute denotes class attribute of these records in dataset. There are total 44350 number of records which is 32.5% of 20071010.txt. The statics of the dataset is as follows:

Table 1: Types of Records

S.N.	Data type	Data Class
1	Unknown Attacks	-2
2	Known Attacks	-1
3	Normal	1

3. Result

This sub section analysis the proposed work on the following Parameters:

3.1 Accuracy

Accuracy is a parameter that is used to find the efficiency of the work. This will show the effectiveness of the classification of the True Positive and True Negative.

Table 2: Accuracy of the Proposed work with Existing Work

Measuring Parameters	Existing Work	Proposed Work
Accuracy	0.724185	0.928508

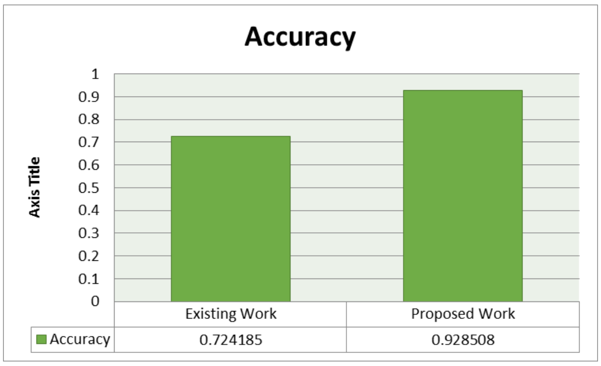

Figure 5: Accuracy of the Proposed work with Existing Work

Table 2 with figure 4 clearly shows that the proposed work is more effective over the existing works. These table and figure clearly show that the accuracy to find accurate record is fairly higher in proposed work.

3.2 True Positive Rate

This is another statics quantity which plays a critical and vital role in finding the efficiency of the work. This quantity shows the efficiency of the work in finding the number of true results out of all possible positive records.

Table 3: True Positive Rate of Proposed work with Existing Work

Measuring Parameters	Existing Work	Proposed Work
True Positive Rate	0.604765	0.666667

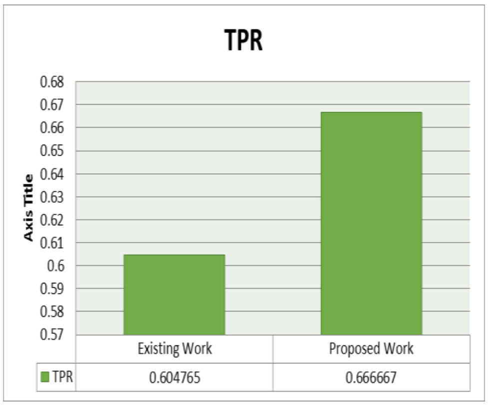

Figure 6: True Positive Rate of Proposed work with Existing Work

Table 3 with figure 5 clearly shows that the efficiency of the proposed work over the existing works. This value clearly shows that the finding capacity of the positive record is fairly higher in proposed work.

7. CONCLUSION

In this paper suggested the Intrusion detection is still a fledging field of research. This field is still in infancy mode. IDS users depend on the IDS to protect their computers and networks demand that IDS provides reliable and continuous detection service. However, other kinds of preprocessing techniques and data mining approach like AI,NN schemes may be tested for a better detection rate in the future research in IDS System. In this paper have done analysis for detecting the malicious misbehaviors. Proposed work is doing more efficiently than the existing work.

8. REFERENCES

[1] Lata, InduKashyap, " Study and Analysis of Network based Intrusion Detection System", International Journal of Advanced Research in Computer and Communication Engineering Vol. 2, Issue 5, May 2013

[2] N.S.CHANDOLIKAR, V.D.NANDAVADEKAR, "Comparative Analysis of two Algorithms for intrusion attack classification using KDDCUP DataSet", International Journal of Computer Science and Engineering (IJCSE) Vol.1, Issue 1 Aug 2012.

[3] Devikrishna K, Ramakrishna B, "An Artificial Neural Network based Intrusion Detection System and Classification of Attacks ", International Journal of Scientific & Engineering Research, Volume 6, Issue 1, January-2015 1370

[4] Vaishali Kosamkar, Sangita S Chaudhari," Improved Intrusion Detection System using C4.5 Decision Tree and Support Vector Machine",International Journal of Computer Science and Information Technologies, Vol. 5 (2) , 2014.

[5] Sisily Sibichen and Sreela Sreedhar,2013. The study of An Efficient AODV Protocol and Encryption Machanism for Security Issues in Adhoc Networks. In Proceeding of IEEE (International Conference on Microelectronics,Communication and Renewable Energy.

[6] P.Kavitha and RajeswariMukesh, 2015. The study of To Detect Malicious Nodes in the Mobile Ad-Hoc Networks using Soft Computing Technique. In Proceeding of IEEE (IEEE SPONSORED SECOND INTERNATIONAL CONFRENCE ON ELECTRONICS AND COMMUNICATION SYSTEM .

[7] Rajendra V. Boppana, Senior Member, IEEE, and Xu Su, Member, IEEE A Distributed ID for Ad Hoc Networks, 26th International Conference on Advanced Information Networking and Applications 2012.

[8] Leila Mechtri, Fatiha Djemili Tolba, Salim Ghanemi, MASID,"Multi agent based intrusion detection in MANET", IEEE 2012.

[9] Monita waghengbam and ningrila marchang,"Intrusion detection in MANET using fuzzy logic", IEEE 2012.

[10] X. Zhang, L. Jia, H. Shi, Z. Tang and X. Wang, "The Application of Machine Learning Methods to Intrusion Detection", Engineering and Technology (S-CET), 2012 Spring Congress on, (2012), pp. 1-4.

[11] Robert Mitchell and Ing-Ray Chen, "Behavior Rule Specification-based Intrusion Detection for Safety Critical Medical Cyber Physical Systems", IEEE Transactions on Dependable and Secure Computing (Volume:12 , Issue: 1), 2014

[12] Devikrishna K S and Ramakrishna B B, "An Artificial Neural Network based Intrusion Detection System and Classification of Attacks", International Journal of Engineering Research and Applications (IJERA) ISSN: 2248-9622, Vol. 3, Issue 4, Jul-Aug 2013, pp. 1959-1964

[13] Vaishali Kosamkar, Sangita S Chaudhari, "Improved Intrusion Detection System using C4.5 Decision Tree and Support Vector Machine", (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (2) , 2014, 1463-1467

[14] Emma Ireland et al. 2013. Intrusion Detection with Genetic Algorithms and Fuzzy Logic. UMM CSci Senior Seminar Conference, Morris, MN.

[15] Ahmed A. Elngar et al. 2013. A Real-Time Anomaly Network Intrusion Detection System with High Accuracy. Information Science Letters.

[16] A.M. Chandrasekhar, "Intrusion Detection Technique By Using K-Means, Fuzzy Neural And Svm Classifier ", 2013 International Conference on Computer Communication and Informatics (ICCCI - 2013), Jan 04-06, 2013 Coimbatore, India.

[17] Elhadi M. Shakshuki, Nan Kang, and Tarek R. Sheltami, "EAACK- A Secure Intrusion Detection System for MANETs" IEEE trans. Vol.60, no.3, MAR, 2013.