An Intrusion Detection System using KNN-ACO Algorithm

Satyendra Vishwakarma
Computer Science & Engineering
Samrat Ashok Technology Institute
Vidisha, (M.P.), India

Vivek Sharma
Computer Science & Engineering
Samrat Ashok Technology Institute
Vidisha, (M.P.), India

Ankita Tiwari
Computer Science & Engineering
Sagar Institute of Research and Technology
Bhopal, (M.P.), India

ABSTRACT
With the remarkable enlargement of the usage of computers through the network and expansion in application running on several platform captures the consideration toward network security. This hypothesis exploits security susceptibilities on the entire computer systems that are technically challenging and expensive to resolve. Therefore, intrusion is employs as a key to conciliate reliability, availability and privacy/confidentiality of a computer resource. An Intrusion Detection System (IDS) participates a noteworthy responsibility in detecting anomalies and attacks over's network. In this research work, data mining conception is integrated with IDS to sort assured the relevant, concealed information of interest for the user efficiently and with fewer implementation times. Four concerns likely Classification of Data, Lack of Labeled Data, Extreme Level of Human Interaction and Effectiveness of D-DOS are being resolved by using the projected algorithms like EDADT algorithm, Semi-Supervised Approach, Hybrid IDS model and transforming HOPERAA Algorithm respectively. In this paper, proposes a SVM and KNN-ACO method for the intrusion detection and the analysis of this is perform using KDD1999 Cup dataset. This proposed algorithm shows improved precision and concentrated false alarm rate when matched with existing algorithms.

Keywords
Precision, Data Mining, Intruders, MATLAB, KDDCUP'99 Dataset

1. INTRODUCTION
The hasty progress and gaining popularity of internet is resulted to the security of networks is increasingly become great connotation and it has been a focus in the current research. The use of internet continuously expanding with an exponential swift, accordingly also is cyber-attacks by hackers fabricating errors in internet protocols (IP), operating system (OS) and application software (AS). A number of shielding methods such as firewall have been situate in place to indorse the functions of intruders which could not assurance the complete fortification of the system. The consequent, the need for a supplementary self-motivated mechanism which resembling to intrusion detection system (IDS) when a subsequent line of resistance.

Intrusion detection is the mechanisms of supervising the events happening in a network and scrutinizes them for indications of intrusion. This system can be alienated into two sorts of used to classify intrusion namely anomaly (AIDS) and misuse intrusion detection system (MIDS) [1-3]. An anomaly exposure mechanism fashions the report of normal activities of uses, system resources, operating system, network traffic and amenities with the audit trails generate by a host or network operating system and network scrutinizing program. This system distinguishes intrusion by identifying noteworthy deviations for the customary behavior patterns of these contours. Anomaly detection system strength is that preceding conception of the sanctuary contravenes of the objective systems is not requisite. For that reason, it is able to perceive not solitary know intrusions nevertheless also unknown intrusions. Additionally, this system can notice the intrusions that masquerade with no breaking safekeeping guidelines [4-5]. The downsides of this mechanism were it had elevated false positive detection error, the involvedness of handling plodding misbehavior, and affluent computation. In this work, we recommend a data mining mechanism for the exposure of network intrusion and the analysis of this approach is done using KDDCUP'99 dataset.

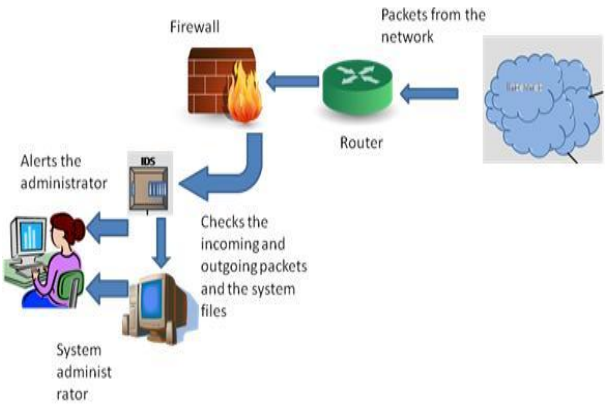

Fig. 1.1: Architecture of Intrusion Detection System

The organization of the remaining section of research paper is done as follows: Section II presents the former work done by various researchers for detection of intrusion. Section III describes the KDDCUP'99 test dataset. In section IV describes our proposed approach to discover the novel threat which compromised from the networks. Section V shows the experimental results and comparative analysis between propose and existing system. Last section presents the overall conclusion of our propose approach which is more efficient than existing.

2. RELATED WORK
This section presents the previous work for intrusion detection by various researchers.

Shona and Senthilkumar [18] In this projected weighted minkowski based firefly algorithm is implemented to abolish redundant record set and improved KNN based allegation method with the help of bagging technique to control missing value is introduced. The there is more enhancement in the precision of learning algorithm throughout classification of

normal and abnormal packets.

Divyamika et al. [6] presented a novel approach to fabricate a network based intrusion detection system using machine learning approach. They have projected two-tier architecture to perceive intrusions on network level. The network behavior can be classified as misuse detection and anomaly detection. As their analysis depend on the network behavior. They have considered data packets of TCP/IP as our input data. After preprocessing the data by parameter filtering, they construct an autonomous model on training set using hierarchical agglomerative clustering. Further, data obtains classified regular traffic pattern or using IDS using KNN classification. They lessen cost overhead. Misuse detection conducted using MLP algorithm. The anomaly detection system is accomplished using Reinforcement algorithm where network agents learn from the environment and take decisions accordingly. The TP rate of our architecture is 0.99 and false positive rate is 0.01. Therefore, their architecture makes available a high level of security by providing high TP and low false positive rate. And, it also analyzes the usual network patterns and find outs incrementally (to fabricate autonomous system) to detach normal data and threats.

Laskov et al. [7] developed an investigational framework for proportional analysis of supervised (classification) and unsupervised learning (clustering) systems for detecting malevolent activities. They used two scenarios to evaluate the learning from both categories. They took training and test data from the similar unknown distribution. The subsequent scenario is based on the novel or unseen data/patterns. This helps us to comprehend how many an IDS can generalize its knowledge to new malicious patterns, which is often very essential for an IDS system.

Wankhede et al. [8] presented an association rule mining technique for IDS. The association mining was applied in order to produce the frequent patterns for different known attacks. These frequent patterns provide a baseline for preventing the attacks from entering the system and also distinguish attacks from normal data, thus allowing normal data to enter the system.

Panda et.al, [9] anticipated hybrid intelligent resolution technologies using data filtering by including guided learning mechanism along with a classifier to make more classified decisions in order to perceive network attacks. It is observed from the consequence obtained that the Naive Bayes model is moderately appealing because of its truthfulness, elegance, robustness and effectiveness. On the other hand, decision trees have demonstrated their effectiveness in both generalization and detection of novel attacks. The results showed that there is no solitary best algorithm to outperform others in all circumstances. In certain cases there might be dependence on the characteristics of the data. To prefer an appropriate algorithm, a domain expert or expert system may employ the outcomes of the classification in order to make better decisions.

Hemalatha et al. [10] introduced data mining perception is integrated with IDS to recognize the relevant, concealed data of interest for the customer efficiently and with fewer execution time. Four problems likely Categorization of Data, High Level of Human Interaction, Short of Labeled Data, and expediency of Distributed Denial of Service Attack (DDOS) are being solved using the projected algorithms similar to EDADT algorithm, Hybrid IDS model, and Semi-Supervised technique and altering HOPERAA Algorithm respectively. Their projected algorithm has been tested using KDD Cup dataset. The entire projected algorithm showed better precision and reduced false alarm rate when compared with already existing algorithms.

Hassanat et al.[11] introduced a novel dataset is collected because there were no widespread data sets that enclose modern DDoS attacks in disparate network layers, such as (SIDDoS, HTTP Flood). This work integrated three well-known categorization mechanisms likewise Multilayer Perceptron (MLP), Random Forest and Naïve Bayes. The experimental outcomes showed that MLP achieved the maximum precision rate (98.63%).

Norouzian et al. [12] employed a most effective classification approach for identifying and classifying attacks into two classes normal or threat. They anticipated a new approach to IDS based on a Multi-Layer Perceptron Neural Network to recognize and categorize data into 6 groups. They implemented their MLP intended with two hidden layers of neurons and achieved 90.78% accuracy rate.

Dong et al. [13] researched the intrusion detection dilemma of the network defense, indicating at the intricacy of low fitting defects in the conventional revealing algorithm of utmost precision and less forecasting exactness under the circumstances of minuscule sample training, and puts forward the algorithm of Support Vector Machine (SVM). Aimed at the considerable influence of SVM kernel function on classification performance, they adopted an improved Ant Colony Algorithm as the procedure of assortment SVM characteristics parameters. Experimental analysis showed this algorithm is appreciably privileged than the other algorithm in training and the exposure speed, and have a better progress of the detection rates of attacking sample.

Barakat et al. [14] novel feature selection model is projected; this model can effectively opted the highly relevant features for intrusion detection. Their aspiration was to build a lightweight intrusion detection system by using a concentrated features set. Deleting irrelevant and superfluous features helps to fabricate a faster training and testing process to have a fewer resource consumption as well as to sustain utmost detection rates. The effectiveness and the practicability of their feature selection model were demonstrated by numerous analysis performed on KDD intrusion detection dataset. The experimental end results vigorously showed that their model is not merely able to capitulate greater detection rates but also to speed up the recognition process.

Tiwari and Rathore [17] in this work their exploration is carried out with respect to two noteworthy assessment metrics such as True Positive (TP)/Recall and Precision/correctness for an Intrusion Detection System (IDS) in KDD cup 99 dataset. Since an outcome of this experiential exploration on the KDDcup'99 dataset, the involvement of all of four assault classes of attributes on Recall and Precision is exemplify which can assist to progress the correctness of KDD cup 99 dataset which accomplish highest precision with lowest false positive (FP).

3. KDDCUP'99 DATASET

Within the year of 1998 the Defense Advanced Research Projects Agency (DARPA) intrusion exposure estimation created the primary standard corpus for estimating intrusion detection systems. The offline intrusion revealing 1998 evaluation was the most important in an intended series of yearly appraisal accomplished by the Massachusetts Institute of Technology Lincoln Laboratories under DARPA defense. For scheming together forged alarm rates and revealing rates of intrusion exposure systems was premeditated by corpus using quite a few types of in cooperation with recognized and novel

attacks enclosed in a huge sum of ordinary surroundings traffic. More than 300 attacks were incorporated in the 9 weeks of data composed for the assessment. These 300 attacks were anxious from 32 distinct attack types and 7 disparate attack scenarios as publicized in KDD dataset. Preliminary observations of the evaluation outcomes for the 1998 competition accomplished that most IDSs can fundamentally identify older, recognized attacks with a little false-alarm rate, even though do not executed as well when identifying novel or fresh attacks. A plentiful extra intrusion detection challenges, similarly DARPA 1999 and KDD Cup 1999, used associated data sets to determine outcomes in intrusion detection exploration. The DARPA 1999 estimation used a related organization for the antagonism, but integrated Windows NT workstations in the simulation network. These assessments of developing technologies are indispensable to focus endeavor manuscript existing prospective and guide investigation. The DARPA assessment focused on the advancement of evaluation corpora that could be used by numerous researchers for system designs and enhancement. The measurement used for the Receiver Operating Characteristic (ROC) system to approximate intrusion detection systems. The ROC method analyzed the tradeoffs amid between false alarm rates and detection rates for detection method. ROC curves for intrusion exposure designate how the detection rate transforms as internal thresholds are varied to engender fewer more or smaller amounts of false alarms to tradeoffs exposure precision beside analyst workload. The training data was connecting to four gigabytes of squashed binary transmission control protocol set out of data from seven weeks of network traffic which was practiced into concerning five millions of group records. In addition to two week of test data authority approximately two million organized record. Accomplishing focused on the systems potential to perceive novel attacks in the test data that was a divergence of an acknowledged attack labeled in the training data. The KDD 99 training datasets enclosed an overall of 24 training attack categories, with an accompanying 14 attack class in the test data merely [15]. The contributors were given a directory of high-level features that could be used to tell apart ordinary relations from attacks. An association is a progression of TCP packets starting and finishing at an simply some well-defined times, surrounded by which data stream from a source IP address to a objective IP address under simply some well-defined protocols. Every association is labeled as either usual or as an attack with precisely one unequivocal attack type. All company manifestation incorporated of approximately 100 bytes. Three sets of attribute were made accessible for analysis. Initially, the similar host characteristics scrutinize merely the associations in the precedent two seconds that have the indistinguishable destination host as the contemporary link, and finish statistics related to protocol activities, provision, etc. The analogous identical service features inspect individual the relations in the last two seconds which have the indistinguishable service as the existing link. The indistinguishable host and undistinguishable service characteristics are mutually called time based traffic features of the link records. Various probing attacks scrutinize the hosts using a much more time period than two seconds, e.g. one time per minute. Consequently, supplementary features can be assembled using a window of 100 links to the similar host as an unusual of a time window. This accepts a group of so called host based traffic features. Ultimately, domain knowledge can be used to engender features that come out for apprehensive behavior in the data segment such as the amount of failed login effort. These characteristics are called content characteristics. The networking attacks fall into four major categories [16]:

A. Denial of Service Attack (DOS): It is the class of attack in which the attackers compile some calculating or memory resource a lot of more eventful or much packed to control justifiable requests, or contradict legitimate users access to a machine.

B. Remote to Local Attack (R2L): It take place simply when an attacker who has the endowment to convey packets to a machine over a network other than who does not have an explanation on that machine extend some susceptibility to add local admittance as a user of that machine.

C. Probing Attack: Such type of attack takes place to gather information concerning a network of computers for the noticeable principle of circumventing its safekeeping controls.

D. User to Root Attack (U2R): It is a class of attack in which the attacker begins out with admittance to a ordinary user account on the system and is dexterous to widen some susceptibility to gain root access to the system.

Table 3.1: Details of Attacks of Labeled Records

Category of Attack	**Attack Name**
Denial of Service	Neptune, Smurf, Pod, Teardrop, Land, Back
Probe	Port_sweep, IP_sweep, N_map, Satan
U2R	Buffer overflow, LoadModule, Perl, Rootkit
R2L	Guesspassword,Ftp_write,I_map,Phf,Multi_hop,Warezmaster,Warezclient

4. PROPOSED WORK

In the categorization of big data domains, occasionally concealed data possibility has been happen as the categorization process. Therefore generated features enclose the false correlations which are not up to the mark of finding the method of intrusion detection. The weakness of extra features is that it restrains massive time for the process of computing and it blows the precisions of IDS. Now feature selection advances the more categorization precision by searching for the suitable features, which greatest classifies the training data. Consequently in the proposed system probability has been premeditated of the each autonomously attributes, afterward entropy has been deliberated and finally information gain has been calculated for each every attributes disconnectedly. And here they applied some logical implies that if calculated gain is incredibly less after that type of attribute will not be contributed for the data preprocessing. Hence, in conclusion 14 attributes found whose gain is higher and that development is done in feature extraction and feature reduction.

The entropy and gain of the proposed work is estimated as follows:

Entropy: It is represented by H(X)

$$H(X) = -P_1 \log_2 P_1 - P_2 \log_2 P_2 - P_3 \log_2 P_3 \dots\dots\dots\dots - P_n \log_2 P_n$$

$$= -\sum_{i=1}^{m} P_i \log_2 P_i \quad \text{.....(1)}$$

The equation 1, the category-wise probability has been

developed after that entropy has been calculated of every independent attributes.

Subsequently gain is calculated as follows:

$$Gain = H(X) - H(X|Y) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (2)$$

Proposed Algorithm

Initial: Read All 41 features of KDD_dataset

Step 1: Apply partial ID3 algorithm for calculate information gain.

Step 2: On the bases of maximum gain selects 14 features in kdd dataset.

Step 3: Read reduced feature dataset from Class data file. % having different classes of attacks and normal

Step 4: Initialize population of Ant's, than they generate pheromone.

Step 5: Position the intensity of pheromone check if related with some specified feature.

Step6: iteration characterize if feasible, it can be dynamic

Step7: Utilize KNN as classifier for testing of all five data class which is classified or misclassified.

Step 8: In data fraction divided dataset, for training and testing training process is done as follows: [train, test] = crossvalind ('holdout', groups);

train1 (: k2-1) = train;%store class-wise training data (:,dos/probe/u2r/r2l) in separate variable

test1(:, k2-1) = test; %store class-wise test data (:, dos/probe/u2r/r2l) in separate variable

Step9: Whichever ant randomly is allocated to one feature and it should visit all five features and amalgamate the subclasses into their parent class.

Step10: Evaluation of the preferred sub classes performed next organize identified sub classes according to classifier performance and distance, after that choose the best optimized categories.

Step11: Verify the criterion status, if the number of iterations is more than the maximum permissible iteration,

Terminate here,

Else

Continue

Step12: Updating pheromone. Lastly, consent to the best ant to deposit an additional pheromone (classes) on parent class.

Step13: Production of novel ants %here earlier ants are removed and novel ants will be produced for current selected feature

Step14: Enhance counter of identified classes

Repeat from step2 to 11 until iteration is not finished

Step15: Classification is done by knn

class = knnclassify(A, A(train1(:, i1),:), groups(train1(:,i1)), k, 'euclidean', 'nearest');

Step 16: After classification process the accuracy of whole progression is calculated for separate class:

cp = classperf(class,groups);

acc = 100*cp.Sensitivity;

5. KDDCUP'99 DATASET

KDD99cup data set used for the principle of experimental research analysis, When they identifies that KDD 99 dataset has been extensively used for the appraisal of signature based intrusion detection. In the novel method they have used KDDCup'99 intrusion detection dataset, which comprises 26167 records with.50:50 training ratio.

Classes of attack types:

1. DOS (Denial of Service)
2. R2L (Remote to Local)
3. U2R (User to Root)
4. Probe

5.1 Set Up of GUI Environment

The projected IDS have been executed in MATLAB2012A [17] tool and the system configuration is Intel I3 core 2.20 GHz processor with 4GB RAM, windows 7 home basic. The projected methodology have primary used the partially ID3 algorithm for the feature reduction from the KDD, the SVM train function is use for training reason of the trained sample, after that KNN is employ for the clustering and classification process for the categorize or miss-categorization of the data, where ACO is ensemble with KNN to improve the best classification rate and optimized the result in very proficient manner. Now classification has five classes' data which is (normal, dos, u2r, r2l, and probe). The following are the lists of features used to perceive the viruses in KDD Cup 1999 dataset. The 41 features are listed in the website [18].

KNN-ACO classified data which were misclassified by only SVM and KNN then applying KNN-ACO on multiple classifiers. This method is emphasis on misclassified classifiers. Where ACO is putted extra efforts to minimize best classified of the category until they are not precisely classified. Subsequent method has been tested on complete (41 attributes) dataset as well as in reduced dataset (18 attributes), and used dimension parameters are: False alarm rate and accuracy and method is compared with SVM, it is found that anticipated method produced most precise outcome into maximum cases.

5.2 Result Analysis

The comparative analysis of the anticipated method KNN classifier with Ant Colony Optimization (ACO) and preceding method Back propagation neural network and Support Vector Machine (SVM) is done using well known performance measuring parameter which is accuracy and false alarm rate. Here, table 5.1 shows accuracy result of the proposed method and existing method. After analysis it is found that the overall accuracy rate for proposed method is about 94.17% whereas the previous exiting method is Back Propagation and Support Vector Machine there accuracy is 93% and 83%. So it is concluded that our proposed method generates more accurate result for intrusion detection as compare to previous method. The result of these method is depicted through graph is shown in fig. 5.1.

Table 5.1 Comparison for accuracy rate of proposed method with BKP, SVM method

Accuracy			
Class of Attack	SVM	BKP	Proposed
DOS	67.24	94.16	92.59

PROB	88.89	92.29	92.21
U2R	89.1	93.97	95.96
R2L	87.15	93.92	95.92

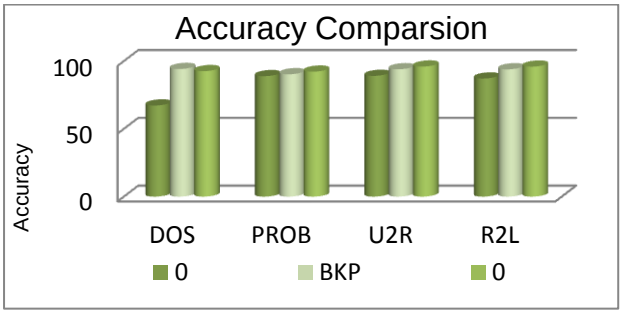

Fig. 5.1: Comparison for accuracy rate of proposed, BKP and SVM method

Here, table 5.2 also shows the false alarm rate result of the proposed method and existing method. After analysis it is found that the overall false alarm rate of proposed method about 5.82% while for other previous exiting method is for Back Propagation is 6.90% and for Support Vector Machine is 16.90% whereas the false alarm rate of proposed method is less than previous exiting method. So it is concluded that our proposed method generates low false alarm rate result for intrusion detection.

Table 5.2: Comparison for overall Accuracy rate and overall FAR

FAR			
Class of Attack	**SVM**	**BKP**	**Proposed**
DOS	32.76	5.84	7.4
PROB	11.11	9.71	7.79
U2R	10.9	6.03	4.04
R2L	12.85	6.08	4.08

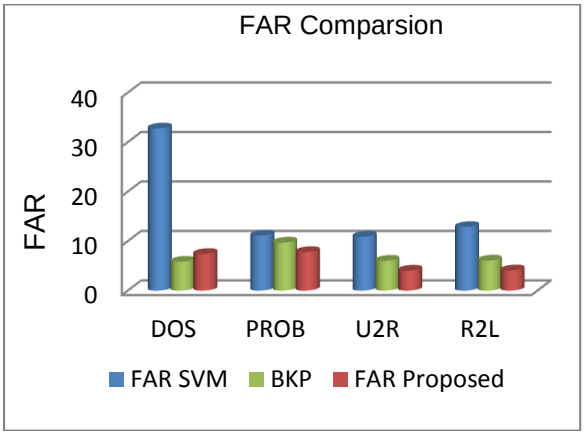

Fig. 5.3: Comparison for False Alarm rate of proposed, SVM and back propagation method

Table 5.3: Comparison for overall Accuracy rate and overall FAR

	Proposed	**BKP**	**SVM**
Overall Accuracy	**94.17%**	**93.58%**	**83.09%**
Overall FAR	**5.82%**	**6.90%**	**16.90%**

6. CONCLUSION

Intrusions detection system is the way of analyzing the traffic in order that the superfluous packets that may surround virus or harm to the network can be detected and countermeasure. The presence of missing values in a KDD cup 99 dataset can influence the performance of a classifier developed using that dataset as a training sample. In this we proposes a SVM and KNN-ACO based methodology to improve the intrusion detection system and after analysis it is found that the performance of the proposed method has significantly improved the classification accuracy and thus it reveals the importance of preprocessing in IDS.

7. REFERENCES

[1] Axelsson, S., "Intrusion Detection Systems: A Taxonomy and Survey," Technical Report No 99-15, Dept. of Computer Engineering, Chalmers University of Technology, Sweden, March 2000.

[2] Lunt, T. F., "Detecting Intruders in Computer Systems," in proceeding of 1993 Conference on Auditing and Computer Technology, 1993.

[3] Sundaram, A. "An Introduction to Intrusion Detection," The ACM Student Magazine, Vol.2, No.4, April 1996. http://www.acm.org/crossroads/xrds2-4/xrds2-4.html.

[4] Porras, P. A., "STAT: A State Transition Analysis Tool for Intrusion Detection," MSc Thesis, Department of Computer Science, University of California Santa Babara, 1992

[5] Dorothy E. Denning, "An Intrusion Detection Model," In IEEE Transactions on Software Engineering, Vol.SE 13, Number 2, page 222-232, February 1987.

[6] Divyatmika, Manasa Sreekesh, "A Two-tier Network based Intrusion Detection System Architecture using Machine Learning Approach", International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT) – 2016 in proceeding of IEEExplore.

[7] Wenke Lee and Salvatore J. Stolfo "Data mining approaches for intrusion detection", In Proceedings of the7th USENIX Security Symposium - Volume 7, SSYM'98, pages 6–6, Berkeley, CA, USA, 1998.

[8] Rajesh Wankhede, Vikrant Chole, "Intrusion Detection System using Classification Technique", International Journal of Computer Applications (0975 – 8887) Volume 139 – No.11, April 2016.

[9] Mrutyunjaya Panda and Manas Ranjan Patra, "Comparative Study Of Data Mining Algorithms For Network Intrusion Detection" First International Conference on Emerging Trends in Engineering and Technology, pp 504-507, IEEE, 2008.

[10] G.V. Nadiammai, M. Hemalatha, "Effective approach toward Intrusion Detection System using data mining techniques", Egyptian Informatics Journal 2015, in

proceeding Elsevier Pp 37–50.

[11] Mouhammd Alkasassbeh, Ahmad B.A Hassanat, Ghazi Al-Naymat, Mohammad Almseidin, " Detecting Distributed Denial of Service Attacks Using Data Mining Techniques", International Journal of Advanced Computer Science and Applications, Vol. 7, No. 1, 2016.

[12] M. R. Norouzian and S. Merati, "Classifying attacks in a network intrusion detection system based on artificial neural networks," in Advanced Communication Technology (ICACT), 2011 13th International Conference on, pp. 868–873, IEEE, 2011.

[13] JianfengPu, Lizhi Xiao, Yanzhi Li and Xingwen Dong "A Detection Method of Network Intrusion Based onSVM and Ant Colony Algorithm", National Conference on Information Technology and Computer Science (CITCS 2012) Published by Atlantis Press.

[14] Ayman I. Madbouly, Amr M. Gody, Tamer M. Barakat, "Relevant Feature Selection Model Using DataMining for Intrusion Detection System", International Journal of Engineering Trends and Technology (IJETT) – Volume 9 Number 10 - Mar 2014.

[15] S. Selvakani Kandeeban, Dr. R.S. Rajesh : "a genetic algorithm based elucidation for improving intrusion detection through condensed feature set by KDD99 dataset", information and knowledge management ISSN 2224-5758, ISSN 2224-896X Vol. 1, No.1, 2011, www.iiste.org.

[16] Mouaad KEZIH, Mahmoud TAIBI "evaluation effectiveness of intrusion detection system with reduced dimension using data mining classification tools", 2nd International Conference on Systems and Computer Science (ICSCS) Villeneuve d'Ascq, France, August 26-27, 2013; 978-1-4799-2022.

[17] VivekNandanTiwari, Prof. SatyendraRathore, Prof. KailashPatidar "Enhanced Method for Intrusion Detection over KDD Cup 99 Dataset", International Journal of Current Trends in Engineering & Technology, Volume: 02, Issue: 02 (MAR-APR, 2016), ISSN: 2395-3152.

[18] Mrs. D. Shona, M. Senthilkumar "An Ensemble Data Preprocessing Approach for Intrusion Detection System Using variant Firefly and Bk-NN Techniques", International Journal of Applied Engineering Research ISSN 0973-4562 Volume 11, Number 6 (2016) pp 4161-4166.