ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

AN INTELLIGENT NETWORK INTRUSION DETCETION SYSTEM USING DATA MINING AND KNOWLEDGE BASED SYSTEM

[1]ALEBACHEW CHICHE, [2]MILLION MESHESHA
[1]Mizan-Tepi University, Tepi, Department of Information Systems, Ethiopia
[2]Addis Ababa University, Addis Ababa, School of Information Science Ethiopia
E-mail: [1]alebachew.chz@gmail.com, [2]mashe84@gmail.com

ABSTRACT

In this study, an intelligent network intrusion detection and prevention system is presented for detecting network attacks that incorporates a knowledge based system and data mining techniques. To extract hidden knowledge from KDDCup'99 dataset, hybrid data mining process is used. The intrusion dataset for the study is collected from MIT Lincon lab. A predictive model is constructed on total datasets of 63, 661 instances using JRip rule induction, Naïve Bayes,J48 decision tree and Multilayer Perceptron (MLP) Neural Network. During training 99.91% prediction accuracy is achieved by J48 decision tree. So, the J48 model is integrated with knowledge based system automatically for designing intelligent network intrusion detection and prevention system. In addition, knowledge is acquired, represented and organized in the knowledge based so as to suggest possible prevention for detected attacks. Evaluation results show that the proposed system registers 91.43% accuracy in network intrusion detection and 85% in user acceptance testing. This indicates that the performance of the proposed system is promising to design an intelligent network intrusion detection system that can effectively predict and provide a prevention mechanism. The system cannot update the knowledge of prevention techniques automatically which need further researches.

Keywords: *Network Intrusion Detection, Intrusion Prevention, Data Mining, Knowledge Based System*

1. INTRODUCTION

Internet and computers have been utilized by many people and organizations all over the world for undertaking their day-to-day activities. In order to come up with efficiency and up to date issues, most organizations rest their applications and service items on Internet.The network becomes incapacitated for a long duration due to the increase in frequent network based attacks. Such situations cause major loss in the financial sectors of an organisation [1].

On the other hand, network intrusion and information safety problems are ramifications of using internet. In other words, network intrusion is considered as new weapon of world war. Therefore, it has become the general concern of the computer society to detect and prevent intrusions efficiently. The conventional protection strategies like firewall remain static in cases where effective protection is required [2]. The network attack and its counter-attack are monitored, to dynamically protect the network using Intrusion Detection System (IDS). Hence, IDS technique is applied to detect network intrusions [1].

Intrusions are detected in a network or a computer system by monitoring their events to analyse the possibility of intrusions occurring. The signs of intrusions are the violations of security policies, user acceptable policies and any breach in standard practices for security [3]. According to Kumar and Gupta [4], depending on the modelling methods, the techniques used for Intrusion Detection (ID) fall into two major categories: misuse detection and anomaly detection. Misuse detection compares the usage patterns for knowing the techniques of compromising computer security. Anomaly detection, on the other hand, approaches the problem by attempting to find deviations from the established patterns of usage [4]. Based on the sources of data Intrusion detection systems can be Host-based and network-based intrusion detection systems that employ one or both of the intrusion detection methods [5]. Most commercially available Intrusion Detection Systems [6] are signature-base systems. But, it is very difficult to specify a rule for huge network traffic. Therefore, data mining techniques are deployed to overcome the limitations of the rule-based systems intrusion detection systems [4].

Mining the network traffic will be able to read the intrusion patterns to detect any mismanagement, to generate normal activities profile for the detection of anomalies and develop classifiers for the massive amount of audit data to detect network attacks [7].

Although more researchers such as [8],[9],[10],[11],[4],[12] have paid attention to solve network intrusions problem based on KDDCup'99 intrusion dataset, but no one is known about how intelligent network intrusion detection system is developed. However, all those researchers have been constructed a predictive model which can only predict and alert an attack.

A major point of difference between traditional and intelligent network detection is the learning process of the system and the inclusion of knowledge base for preventing network intrusion. Much less researches has investigated the integration of knowledge based system in network intrusion detection systems. The continuing proliferation of network intrusions conveys the need for researches that extends beyond the traditional ways of detecting and preventing network intrusions into a way of detecting and preventing network intrusions intelligently. The need can be described in many ways: For example, all most all of business, educational and governmental institutions are now resting their services on the internet. But they losing their money because of the limited skilled security experts. In addition, although some institutions have security experts, still they are limited in detecting and preventing a new attacks. For example, an organization may have a security experts working on firewall, but they my know few of network intrusions.

Although academic researchers[13],[14], have forward a need for greater understanding in this area, much less is known about how to integrate data mining model with knowledge based systems for network intrusion detection which can affect organization's business success.

That is, this study attempt to integrate data mining model for detecting network intrusions and knowledge based system to come up with an intelligent network intrusion detection system for the first time. Essentially, this study responds to the questions for the new thinking about the integration process of data mining model with knowledge based system and draws inspiration from both Tigabu [13] and Domingos [14], who have stressed the need for network intrusion detection by integrating data mining and knowledge based that will enable system and network administrators both to detect and provide a mechanisms to prevent network intrusions. As defined by Domingos [14], data mining has made tremendous progress these days. However, Domingos have proved that still there is a large gap remains between the results a data mining system can provide and taking actions based on result.

In addition, this study is limited to offline dataset of KDDCup'99 collected from MIT Lincoln Laboratory. And the knowledge based integrated with predictive model is unable to update itself automatically.

The findings of this study are expected to assist network and systems administration practitioners in filling the gap of analysing and preventing network intrusions and increasing awareness to different kinds of novel network intrusions.

According to Chakraborty et,al.[15],a knowledge based system is a computerized system that uses knowledge of a given domain in order to deliver a solution for a problem. As noted by Sajja and Akerkar, Knowledge based systems has been developed for a variety of reasons, including the archiving of rare skills, preserving the knowledge of retiring personnel, support in decision making and to aggregate all of the available knowledge in a specific domain from several experts and/or machines.

In this paper, knowledge engineering approach is followed during the development of knowledge based system for prevention of detected attacks. Knowledge is acquired from domain experts and document analysis and represented using rule based knowledge representation technique.

Therefore, this study aims at integrating data mining with knowledge based system so as to develop an intelligent network intrusion detection system that can alert attacks as efficient and effective as possible and provide a corrective action for network administrators. Hence, the data mining model constructed on KDDCup'99 dataset is used for detecting the attacks, based on which the knowledge based system is enabled to provide possible prevention and corrective action for detected attacks.

2. RELATED WORKS

Research is carried out to design IDS that uses Machine learning algorithms. For this study we have been assessed data mining methods used for network intrusion detection. Since there is no previous literature about the integration of data mining and knowledge based system, here we have tried to present intrusion detection using data mining techniques. Among the works, the framework done by Panda and Patra [8] is based on Naïve Bayes algorithm on the benchmark intrusion dataset- KDDcup'99. The framework builds the patterns of the network services over data sets labeled by the services. Based on time, accuracy and cost evaluation metrics the proposed techniques performed better than back propagation neural network algorithm on KDD data. However, the framework generates more false Positives. Ferid, et al [9], proposed a new learning algorithm for adaptive network intrusion detection using Naive Bayesian classifier and decision tree. The hybrid algorithms perform high detection rates and significantly reduce false positives for different types of network intrusions using limited computational resources. From the result they have concluded that, hybrid algorithm minimized false positives and maximized detection rates on the 5 classes of KDD99 benchmark dataset. Peddabachigari, et, al [10] combined decision tree and support vector machine for detecting network intrusions. The hybrid approach shows that 99% performance on KDDCup'99 intrusion data. They have concluded that hybrid model is better than individual model.

Ankit and Hande [11], applied naïve Bayes for anomaly based network intrusion detection. They have used KDDCup'99 intrusion dataset for testing Naive Bayes classifier. The proposed approach achieved higher detection rate of 97% accuracy on KDDCup'99.

In this study, we present a Network Intrusion Detection System by integrating data mining and knowledge based system for detecting an intrusions and preventing them intelligently. Surveying and comparing with all the works in the literature we proposed intelligent IDS different from the literatures. First we construct an intrusion detection model using data mining algorithms where all the intrusion detection activities are maintained by this intrusion detection model. Following this, we develop a knowledge based system for intrusion prevention based on detection results.

3. ARCHITECTURE DESIGN

To develop an Intelligent Network Intrusion Detection System we propose architecture depicted in figure 1.

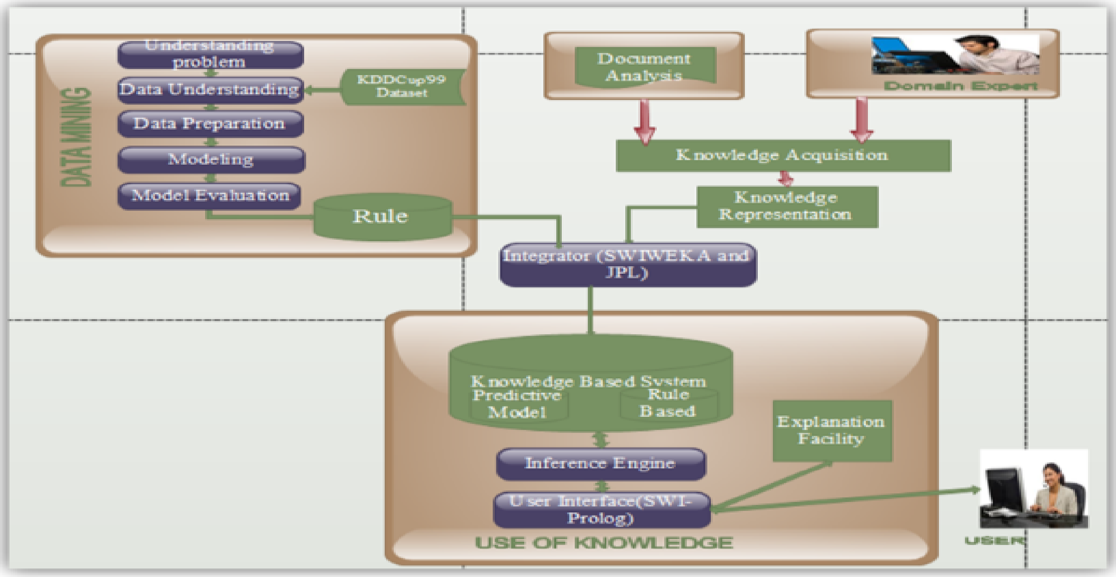

Figure 1: Architecture of the proposed system [16]

As presented in figure 1, the architecture proposed in this paper comprises two major components, namely, intrusion detection and intrusion prevention. For constructing intrusion detection model, data mining technology is used to acquire hidden knowledge about the behavior of the different attacks from the given data. For prevention purpose tacit and explicit knowledge is collected from domain experts and manuals and represented.

Accordingly, for intrusion detection the model constructed by J48 decision tree is used for detecting network attacks. Java programming was used to integrate intrusion detection model created using J48 decision tree with the Knowledge Based System automatically. For this study, we used Weka, jpl and swi-prolog tools to construct a predictive model. Hence, our system automatically takes the data set, construct the model using the selected J48 decision tree classification algorithm and integrate the model with knowledge base for network intrusion detection.

4. CONSTRUCTING PREDICTIVE MODEL

The predictive model is constructed on a KDDCup99 [17] intrusion detection data has been used, which has been collected from Massachusetts Institute of Technology (MIT) Lincoln laboratory. According to [17], those network intrusions labeled as follows: Probing Attack, Denial of Service Attack, Remote-to-Local Attack and User-to-Root Attack.

Denial of Service (DoS) is a type of attack in which network resources busy with an intention to deny genuine user from accessing it. On the other hand, User-to-Root (U2R))in which an attacker uses sniffing passwords or any social engineering to access normal user account then use a vulnerability to gain a root access. The third attack type is Remote-to-Local (R2L) an attack in which a machine that can send packet without account gain local access of user. Finally, probing attack is an attack in which intruders gather information about a network of computers to get general information about a specific network. Hybrid data mining process model is followed for data mining task. We used observation by closely working with domain experts, interviewing experts as well as domain researchers, reviewing documents, reports and literatures to understand the problem domain. After problem understanding, we checked the syntax of the KDDCup'99 dataset, attributes and classes as well as the quality of the content by collecting the dataset from Massachusetts Institute of Technology (MIT) Lincoln laboratory [17].

To prepare the data for data mining techniques, different preprocessing tasks such as data cleaning (for filling missing values, handling duplication and outlier values), dataset reduction and balancing tasks were performed. Data size reduction, dimensionality reduction and balancing the dataset is employed in data preprocessing tasks.

Resampling method has been applied on the KDD dataset to reduce the size of the data and to create a manageable data set. Table 1 shows the preprocessed dataset for all class. This is followed by attribute selection. As stated by Selvakani and Rajiesh [18], feature selection methods are commonly used for cost insensitive learning and found to be very helpful because the elimination of useless features enhances the accuracy of detection while speeding up the computation, thus improving the overall performance of IDS.

Information gain ratio were followed for selecting best attribute selection for data mining that come up 14 out of 41 attributes. These attributes are urgent, dst_bytes, logged_in, protocol_type, lnum_file, is_guest_login, wrong_fragment, is_host_login, dst_host_srv_diff_host_rate, service, rerror_rate, count, creations, srv_diff_host_rate, and duration.

Balancing is needed if one target class has much lower frequency than the other target class in the given dataset [19]. Because over fitting is likely to occur whenever one class has a large number of possible values [20]. SMOTE (Synthetic Minority Over-sampling Technique) is an approach that over-sampled minority class and introduces synthetic examples along the line segments joining any/all of the k minority class nearest neighbors [21].

As shown in Table 1, the classes from sample imbalance dataset were balanced using SMOTE technique.

Table 1: Number of records after preprocessing

Classes of Intrusion	Before SMOTE		After SMOTE	
	Number of records in each class	Share of each class in percent (%)	Number of records in each class	Share of each class in percent (%)
Normal	21,352	52.99	21,352	33.54
DOS	18,469	45.23	18,469	29.01
Probe	626	1.53	10,016	15.73
U2R	7	0.22	7,168	11.26
R2L	104	0.028	6656	10.46
Total	40,558	100	63,661	33.54

The experiments were conducted with four classifiers: namely algorithm JRip using induction rules, Decision tree with J48, Multi-layer Perceptron and Naive Bayes. The data set has five classes such as Normal, DoS, probe, U2R and

R2L.Experimental results using 10-fold cross-validation is presented in table 2 below.

Table 2: Summary of experimental results

Algorithms	Accuracy (%)	Weighted average precision (%)	Weighted average recall (%)	Weighted average F-measure (%)
J48 decision tree	99.91	99.9	99.9	99.9
JRip rule induction	99.89	99.9	99.9	99.9
Naïve Bayes	67.69	76.1	67.7	64.1
Multi-layer Perceptron (MLP)	98.01	98	98	98

As we can observe from the table above, J48 decision tree outperforms other classification algorithms, with an accuracy of 99.91 percent. On the other hand, J48 decision tree scores 99.9 percent on precision, which is coequal with JRip rule induction. The third is recall; in this measure J48 decision tree outperforms Naïve Bayes and MLP neural network classification algorithms with 99.9 percent. Finally, F-measure has been seen for comparison, in this measure J48 decision tree outperforms with 99.9 percent which is comparable with JRip rule induction classification algorithm.

Based on objective evaluation results, the network intrusion detection model constructed by J48 decision tree is selected as the best working model for this study. The confusion matrix for the selected model is presented in table 3 below.

Table 3: Confusion matrix for J48 decision tree algorithm

Predicted classes					
Normal	DOS	Probe	U2R	R2L	Actual classes
21331	7	8	1	5	Normal
10	18457	1	0	1	DOS
9	0	10,007	0	0	Probe
1	0	0	7165	2	U2R
11	0	0	2	6643	R2L

As mentioned earlier, J48 decision tree, JRip rule induction, Naïve Bayes and MLP neural network algorithms are adapted for the experiments. The results of the algorithms are evaluated based on prediction accuracy in classifying the instances of the data set into normal, probe, DOS, R2L and U2R. As indicated in the table 2, the classifiers performed almost the same. There is a slight difference among the classifiers in terms of classifying the data set correctly.

Even though their slight difference, J48 has registered the best prediction accuracy by classifying 40,494 instances out of 40,558 correctly. Results of JRip and J48 show that nearly equal number of incorrectly classified instances. The highest incorrect classification is registered by Naïve Bayes algorithm.

Table 3, depicts the confusion matrix for best performing classifier. So, from the confusion matrix one can understand that selected model classified 21,373 instances as normal but expected to classify only 21,352 instances. The other 6, 8, 1 and 1 instances were misclassified as DOS, probe, U2R and R2L actually they are normal. The selected model classified 18,470 instances as Denial of Service attack but expected to classify only 18,469 instances and incorrectly classified 7 instances as normal, 1 instance as probe. And also classified 610 instances as probe but it is expected to classify 626 instances as probe and incorrectly classify 13 instances as normal and 3 instances as DOS attacks. The model also classified 4 instances as User to root attacks but expected to classify 7 instances as user to root attacks and incorrectly classified 3 instances as a normal. It also classified 84 instances as root to local attack but expected to classify 104 instances as root to local attack and incorrectly classified 18 instances as normal, 2 instances as user to root attack.

In this study, we have used prediction accuracy to select best working model. Based on this, the comparative analysis of the models shows that the J48 decision tree classifier outperformed best with a classification accuracy of 99.91 % . The JRip classifier came to be second best with a classification accuracy 99.82% And the Naïve Bayes classifier came out to be the worst with classification accuracy of 92.86%, Empirical results depict that the decision tree J48 classifier gives better performance for detecting attacks than all other three individual algorithms. After all we have selected decision tree J48 classifier as a better performed model to integrate with knowledge base.

The reason for the J48 decision tree performing better than Naïve Bayes as well as MLP is because of the linearity nature of the dataset and decision tree is a straight forward in it nature. This means there is a comprehensible segregation point that can be defined by the algorithm to predict the class of a particular network intrusion. The other reason for the Naïve Bayes and MLP neural network scoring a lower accuracy than the J48 decision tree is because both algorithms need complex data, therefore loss of accuracy. In addition, in terms of ease to interpret and implement the J48 decision tree is more self-explanatory. It can handle large number of features and generates rules that can be converted to simple and easy to understand classification if-then-else rules.

Therefore, the model which is developed with the J48 decision tree with 10-fold cross validation classification techniques is considered as the selected working model for this study.

Generally, from the confusion matrix (see table 3) one can understand that the selected model classified well the majority of the instances into their proper class. However, there are few cases where intrusions are considered as normal; and also normal instances are classified as intrusion. Such insignificant misclassifications are tolerable and experts can easily identify them during analysis. Hence, the model constructed using J48 decision tree is finally integrated with the knowledge based system so as to come up with an intelligent network intrusion detection system. The algorithm to predict network attack is shown below:

```
Problem: Predicting Network Attacks
Inputs: Attribute values of the network traffic
Output: Predicted network attack
Let userinput, be the attribute value quest by system
Let samplearff, be the arff file created from userinput
Let mainarff, be the training dataset to train algorithm
For I from 1 to N do:
        Read/input userinput from users
       If userinput mismatch attributes value
          Return false
Else Convert userinput to Samplearff
Return samplearff
Read mainarff for training algorithm
Read samplearff to be classified based on training
Set class index for each instances in mainarff, samplearff
Set the selected algorithm X,
For I form 1 to N do
Let C =Train (mainarff, X), return a trained algorithm c
Apply (samplearff, C), return a result R, //where R is a predicted attack
Write R on samplearff as a label and append the labeled samplearff to mainarff // automatically label the instance and append on mainarff
Return mainarff // updated training dataset
```

5. KNOWLEDGE ACQUISITION AND REPRESENTATION FOR INTRUSION PREVENTION

Knowledge acquisition is the process of extracting, structuring and organizing knowledge from human experts and other sources such as books, databases, the Internet, research papers, documents, one's own experience and transferring it to the knowledge base [22].The knowledge for this study is acquired by interviewing domain experts as well as by relevant documents analysis which has been employed to purify the acquired knowledge. After knowledge is acquired, conceptual modeling is used to model the knowledge. For this study, the knowledge acquired from domain experts and manuals are modeled using decision tree to clearly understand the decision making process during the network intrusion prevention.

Once the knowledge has been acquired and modeled, the next step is knowledge representation using appropriate format for this study a rule-based knowledge representation and reasoning is employed that is in line with decision tree for knowledge modeling. This is one of the most commonly used techniques in which knowledge is represented in the form of IF condition-THEN action pairs. Finally, the knowledge is included into knowledge base to use it as courses of action for the detected intrusion. For example:

If detected attack =DOS,

THEN prevention technique =Install a firewall, and configure it to restrict traffic coming into and leaving your computer.

Algorithm for preventing predicted network attack has been discussed as follow:

Problem: Taking action on Predicted Attack
Inputs: Predicted network attack
Output: Possible corrective actions for predicted attack
Algorithm:

```
Let traffic_class, be class of network traffic predicted by model
For each possible value of traffic_class
If traffic_class is DoS attack
        Return list of option with DoS attack.
Else if attack class is probe attack
        Return a list of options with respect to probe attack
Else if traffic_class is U2R attack
        Return a list of options with respect to U2R attack
Else if traffic_class is R2L attack
        Return a list of options with respect to R2L attack
Else return the network traffic is normal
```

6. EVALUATIONS

Two types of testing are done in this study, user acceptance and system performance testing. The latter testing is to verify the system working automatically without the interference of user or expert. 35 test cases were prepared form real network traffic to evaluate the performance of the system. So, for evaluating the system performance confusion matrix was used.

Table 4: Confusion matrix for intelligent network intrusion detection system

Actual class labelled	System predicted value				
	Normal	DoS	Probe	R2L	U2R
Normal	10	0	0	0	0
DoS	0	8	0	0	
probe	1	0	6	0	0
R2L	1	0	0	4	0
U2R	0	0	0	0	5

As a result the proposed system could perform in the absence of domain experts with 91.34% accuracy which indicates that the study was effective in predicting an attack and providing prevention knowledge of detected attack. One of the challenges for the systems during performance testing was that the systems sometimes misclassified R2L and probe Attacks as Normal traffics.

User Acceptance testing: This evaluation method allows expert to make comments while interacting the system. To do so close ended questionnaires were adapted and modified in the context of Intelligent Network Intrusion Detection from Birhanu [23], focusing on easiness, attractiveness, time efficiency, applicability of the system, problem solving ability and the significance of the system in the domain area, accuracy of the system type of questionnaires were distributed for feedback collection from the evaluators.

To evaluate INIDS in this study, domain experts were selected as system evaluators. The experts were selected purposively, who are system administrators. Before starting the evaluation, we explained the objective of the developed system and how the system interacts with the users. This explanation helps the evaluator get full understanding how they consult the system in getting advice.

Then after, the domain experts were allowed to interact with the system by running number of cases having similar parameter with the ARFF file incorporated in the knowledge base. After the consultation of the system, to assess the user acceptance of the system, questionnaires were distributed. Using these questionnaires, domain experts' feedback towards this developed system was gathered for analysis.

The type of questionnaires distributed for feedback collection from the evaluators were closed ended questionnaires focusing on easiness, attractiveness, time efficiency, accuracy of Intelligent Intrusion Detection Knowledge Based System (INIDS). The questionnaires also focused on the applicability of the system in predicting network intrusions, problem solving ability and the significance of the system in the domain area.

The evaluators were allowed to rate the options as excellent, very good, good, fair, and poor for these closed ended questions. Therefore, for easiness of analyzing the relative performance of the prototype based on the user evaluation after the interaction with the system, the researcher assigned numeric value for each of the options given in words.

The values are given as Excellent = 5, Very good = 4, Good = 3, Fair = 2, and Poor = 1. The table below indicates the feedbacks collected from evaluators during systems interaction.

Table 5: Evaluation using user acceptance testing

No	Questions	1	2	3	4	5	Average	%(100)
1	Is the system is easy to use and interact with it?	0	1	0	2	1	3.75	75
2	How do you rate IIDS attractiveness?	0	0	3	0	1	3.5	70
3	Is the system is more efficient in time?	0	0		2	2	4.5	90
4	How accurately does a system reach a decision predicting an attack?	0	0	0	0	4	5.0	100
5	Does the system incorporate sufficient and practical knowledge?	0	0	2	2	0	3.5	70
6	Does the system give right description damages and prevention for predicted attacks?	0	0	0	1	3	4.75	95
7	How do you rate the significance of the system in the domain area?	0	0	0	1	3	4.75	95
		Total average					4.25	**85%**

As indicated in above table 5, 25% of the evaluators replied easiness to use and to interact with the system as fair and the same number of evaluators also rated that easiness to use and to interact with the prototype as excellent. But the highest number (50%) rated easiness to use and to interact with the prototype as very good. In case of attractiveness of the prototype 75% the evaluators replied the prototype as good and the remaining 25% of respondents evaluated the attractiveness of prototype as Excellent. In case of time efficiency, 50% of evaluators have rated it as very good and the remaining 50% of the evaluators rated it as excellent.

Additionally, 100% of the evaluators evaluate the prototype as excellent at deciding accurate decision of an attack and description during prediction of network attacks. In the same way for criteria of the prototype incorporating sufficient knowledge, 50% the evaluator rated as good and the remaining 50% evaluated as very good. Similarly, in terms of the prototype providing right description of damages and prevention for predicted attacks, 25% evaluators evaluated as very good, and 75% of the evaluators replied as excellent.

The final evaluation criterion is importance of the prototype in the domain area. This criteria is included to measure how does INIDS is important in the area of intrusion detection. And 75% of evaluators rated that as excellent and 25% of the evaluators replied as very good. This implies that the contribution of developing INIDS is important in the area of intrusion detection to detect and suggest corrective action for detected network intrusion.

Based on the results obtained the overall average performance of the system with user's point of view is 4.25 on a scale of 5. This result indicates that the overall average performance of the INIDS is about 85%. This implies that the modeled prototype was performs well in making right decisions on the prediction of network intrusions.

Generally, this prototype has got 85% acceptance by experts in predicting network intrusions and providing the corrective action for predicted attack. With time efficiency and cost effectiveness, this prototype Intrusion detection KBS has great value in predicting network intrusions and providing interpretations by using knowledge base stored as rules and facts of network intrusions when consultation about attacks is needed in the areas it can be deployed. Lastly, the system evaluators also provide their suggestions and comments on the weakness and strength of the prototype. According to the system evaluators the following are some of the basic limitations of the knowledge based system. The first one is lack of interactive user interface which can easily understandable by end users, and suggested that it should include graphical user interface. The second issue raised by user is the lack of brief prevention process so suggested to incorporate additional tasks in prevention process. Finally, the user complained in the way the system accepts user input from users, it lacks some flexibility and recommended to use another way of input mechanisms.

As a result, the proposed system achieved 85% of the user acceptance which is the promising result to implement the proposed system. This implies that the prototype perform well in making right decisions on the prediction of network intrusions and provide corrective action for detected attacks

7. CONCLUSION

With the rapid growth of computer networks during the past few years, security has become a crucial issue for modern computer systems. Traditional protection techniques such as user authentication, data encryption, avoiding programming errors and firewalls are used as the first line of defense for computer security. But this classical methods are failed to prevent all network intrusions.

So, in this study we proposed an intelligent intrusion detection system which can predict attacks in the network and suggest the proper corrective actions for predicted attacks. From this work, we conclude that integrating data mining based network intrusion detection with a knowledge based system to come with a knowledge rich intelligent network intrusion detection systems is an ultimate solution for institutions especially which have a limited security expert. Apart from that, now a day's artificial intelligence is becoming a focus for every discipline. The result form the systems shows that it is possible to integrate knowledge base containing a preventive mechanisms for detected network intrusions which helps systems administrators to fill their skill gaps.

The system is developed by integrating data mining model and knowledge based system for detecting intrusion types. A model is constructed to predict the intrusion detection is proposed that uses four classifiers MLP, Naive Bayes, Decision tree using J48 and JRip algorithm using rule induction. Dataset used are samples from MIT Lincoln laboratory were collected. Further, the knowledge for prevention techniques are acquired from domain experts and document analysis. The proposed system achieves 91.34 and 85 percent on system performance testing and user acceptance testing respectively. The result is promising to design an intelligent NIDP system by integrating data mining with knowledge based system. In general, the promising result from this study is realized in integrating machine learning with knowledge based system for detecting network attacks and providing prevention mechanisms to network administrators. In this research the knowledge base was not able to update itself. So, we forward further research towards enabling the system automatically learns prevention mechanisms, conducting further research on local dataset. The current research was limited to supervise learning for constructing predictive model. Future work should examine unsupervised learning. Also future work should be conducted on real time network traffic

REFRENCES:

[1] Chaudhari M, "Review on Data mining techniques for Intrusion Detection System", International Journal of Innovative Research in computer and communication engineering, 2014, vol.2, no.1, pp.2587-2592.

[2] Sagale A, Kale S, "Hybrid Model For Intrusion Detection Using Naive Bayesian And Support Vector Machine", International Journal of Computing and Technology, vol. 1, no.3,2014, pp. 56-59.

[3] Scarfone K, Mell P, " Guide to Intrusion Detection and Prevention Systems (IDPS) Recommendations of the National Institute of Standards and Technology", Nist Special Publication, vol. 800-94, 2007, p. 127.

[4] Kumar P ,Gupta N, "A Hybrid Intrusion Detection System Using Genetic-Neural Network", International Journal of Engineering Research and Applications (IJERA) , March 2014, pp.59-63..

[5] Gujar S, Patil B, "Intrusion detection using Naive Bayes for real time data", International Journal of Advances in Engineering and Technology, vol.7,no.2,2014,pp. 568-574.

[6] Ahmed Y ,Ahmed E, "Network Intrusion Detection using Data mining techniques and network Behavior Analysis", International Journal of Computer Science & Information Technology (IJCSIT), vol.3,no. 6, 2011,pp.87-98.

[7] Denatious D John A, "Survey on data mining techniques to enhance intrusion detection", In 2012 International Conference on Computer Communication and Informatics (ICCCI), Jan 10-12, 2012.

[8] Panda M, Patra M, “Network Intrusion Detection using naive bayes”, IJCSNS International Journal of Computer Science and Network Security, vol.7,no.1, 2007,pp.258-263.

[9] Farid D, Harbi N and Rahman M, “Combining Naive bayes and decision tree for adaptive Intrusion Detection”, International Journal of Network Security & Its Applications,vol. 2,no.2,2010,pp. 12-25.

[10] Peddabachigari S, Abraham A ,Grosan C and Thomas J, “Modeling intrusion detection system using hybrid intelligent systems”, Journal of Network and Computer Applications, vol.30,no.1,2007,pp.144-132.

[11] Anikt p. Sagane, prof.S.S.Dhande, “Malicious Code Detection using Naive bayes classifier”, International journal of application or innovation in engineering, vol.3,no. 4, 2014,pp. 404.

[12] Kosamkar V, Chaudhari S, “Improved *Intrusion* Detection System using C4.5 Decision Tree and Support Vector Machine”, (IJCSIT) International Journal of Computer Science and Information Technologies, vol.5,no.2,2014,pp.1464.

[13] Dagne T. Constructing a predictive model for network intrusion detection [Master’s thesis], Addis Ababa: Addis Ababa University, 2012. 125p.

[14] Domingos P, "Towards Knowledge Rich-data mining," Data Mining and Knowledge Discovery, Springer, vol. 15, no. 1, 2007, pp. 21-28.

[15] Chakraborty S, Roy D and Basu A, “TMRF e-Book Development of Knowledge Based Intelligent Tutoring System”, Advanced Knowledge Based Systems: Model, Applications & Research, Sajja & Akerkar, Eds., Vol.1, 2010, pp.74–100

[16] Chiche A, Meshesha M, “Constructing a Predictive Model for an Intelligent Intrusion Detection”, International Journal of Computer science and Information Security, vol.15,no. 3, march, 2017, PP. 392-397.

[17].MIT Lincoln Lab [Internet]. 1998 DARPA Intrusion Detection Evaluation Program: MIT Lincoln Lab; 2015 [updated 2014 September 20; cited 2015 July 15]. Available from: https://kdd.ics.uci.edu/databases/kddcup99/task.html

[18] Selvakani S, Rajiesh R, “Genetic Algorithm for framing rules for intrusion detection” IJCSNS International Journal of Computer Science and network Security,vol.7,no.11, 2007,pp.285-290.

[19] Larose D T. Discovering Knowledge in Data: An Introduction to Data Mining, New Jersey, USA: John Wiley & Sons Inc.; 2005.239p

[20] Eibe F, Mark A H, and Ian H W. Data Mining Practical Machine Learning Tools and Techniques 3rd ed. USA: Morgan Kaufmann: USA; 2011.

[21] Chawla N Bowyer K Hall L Kegelmeyer W, " SMOTE: Synthetic minority over-sampling technique," Journal of Artificial Intelligence Research 16, 2002, PP. 321–357

[22] Geber M, Mohammed A, “A Hybrid Model for Knowledge Acquisition Using Hierarchical Cluster Analysis” In International Arab Conference of e-Technology, Egypt,2008, pp.1-14.

[23] Aebissa B. Developing a Knowledge based system for coffee disease diagnosis and treatment [Master’s thesis], Addis Ababa: Addis Ababa University, 2012.109p.