Journal of Network and Information Security
7 (1), June 2019, 06-14
http://www.publishingindia.com/jnis

A Combined Reasoning System for Knowledge Based Network Intrusion Detection

Meseret Assefa[1*] and Million Meshesha[2]
[1]Addis Ababa University, Institute of Ethiopian Studies, Ethiopia. Email: meseret.adamu@aau.edu.et
[2]Addis Ababa University, School of Information Science, Ethiopia. Email: million.meshesha@aau.edu.et
*Corresponding Author

Abstract: In this study, a combination of rule based and case based reasoning for network intrusion detection is proposed. To this end, knowledge is extracted using data mining from sampled KDDcup'99 intrusion data set. Both descriptive and predictive models are created using K-means clustering and JRip rule induction respectively. Descriptive model is used to design case-based reasoning and predictive model to construct rule-based reasoning. A conditional combination is used for controlling the reasoning between RBR and CBR. In the combined system, it is the RBR that first treat the new query for recommending a solution. If RBR is unable to recommend, the query is automatically forwarded to the CBR system where the case retrieval module identifies the most related solution using case similarity measure. The combination of rule-based and case-based reasoning methods has shown an average of 9.5% improvement with regards to performance over the individual reasoning methods. As a continuation of the intrusion detection, we are now working towards the development of a combined intrusion detection system that prevents intruders to enhance the performance of the system.

Keywords: Combination of CBR and RBR, Combined intrusion detection, Knowledge-based intrusion detection, Network intrusion detection.

I. Introduction

Computer Security is the protection afforded to an automated information system in order to attain the applicable objectives of preserving the integrity, availability and confidentiality of information system resources such as hardware, software, firmware, information/data, and telecommunications [1].

These days most organizations are dependent on the Internet to communicate with people and systems to provide news, online shopping, email, credit card detail and personal information [2]. Due to the rapid growth in technology and widespread use of the Internet, a lot of problems have been faced to secure the system's critical information within or across the networks because there are millions of people attempting to attack on systems to extract critical information [2] [3].

Usman *et al.* [3] define intrusion as an attempted act of using computer system resources without privileges, causing incidental damage. Intrusion Detection System (IDS) is an approach that monitor network traffic and its suspicious behavior against security. IDS are a set of techniques and methods that are used to detect suspicious activities both at the network and host levels [3].

An IDS is an effective security tool which helps the users and administrators to prevent unauthorized access to network resources. There are various kinds of attacks that are most common in IDS. They are probe attacks, denial-of-service attacks (DoS), remote to local (R2L) and user to root (U2R) attacks [2] [3].

According to Sumit *et al.* [4], IDS perform signature-based monitoring of the cyber infrastructure to identify any malicious activities and generate an alert when such activity is detected. IDS monitor network traffic and its suspicious behavior against security. If it detects any threat then alerts to the system or network administrator about intrusion. Intrusion Detection System can be classified into three categories [5]. The first is Host-Based IDS that evaluates information found on a single or multiple host systems, including contents of operating systems, system and application files. The second is Network-Based IDS, which evaluates information captured from network communications, analyzing the stream of packets traveling across the network. Packets are captured through a set of sensors. The final one is Vulnerability-Assessment IDS, which detects vulnerabilities on internal networks and firewalls [6].

Signature based detection has also disadvantages; Signature-based approaches can only detect misuse for which a signature exists [7] [8]. For a signature to exist, the form of misuse must be known about beforehand so it can be researched and programmatically identified. This means any new form of misuse will not be detected by a signature based system until it

is identified, analyzed, and then incorporated into the product. Depending on the circumstances, this could be hours, days, weeks, or even months [7] [9] [10].

Anomaly Based Detection on the other hand is based on defining the network behavior. Behavior of the network is the predefined one, when it is accepted or else it triggers the event in the anomaly detection. The accepted behavior of the network is prepared or learned by the specifications of the network administrators [10].

According to [11] the important phase in defining the network behavior is the IDS engine that is capable to cut through the various protocols at all levels. The engine is able to process the protocols and understand the goal. Though this protocol analysis is computationally expensive, the benefits it generates like increasing the rule set helps in less false positive alarms.

According to [5] the major drawback of Anomaly Detection is defining its rule set. The efficiency of the system depends on implementation and testing on all protocols that are available. Rule Defining Process is also affected by various protocols used by several vendors.

Intrusions can be detected by observing deviations from the expected behaviors of the system monitored. These "normal" behaviors can either correspond to some observations made in the past or to some forecasts made by various techniques. Everything that does not correspond to this "normal" pattern will be flagged as anomalous [5] [6].

Therefore, the core process of anomaly detection is not to learn what is anomalous but to learn what is normal or expected [7] [11]. The key advantage of anomaly-based detection over the other techniques is the ability to detect new attacks, attacks for which no signature or known protocol violation exists [9].

According to Fayyad *et al.* [6], there are two primary models to analyze events and detect attacks: Misuse detection and Anomaly Detection Models. Misuse detection is an IDS that detect intrusions by looking for activity that corresponds to known signatures of intrusions or vulnerabilities. On the other hand, Anomaly Detection Model is an IDS that detect intrusions by searching abnormal network traffic. Data mining based methods are another paradigm for building intrusion detection system. The main advantage of these methods is that they leverage the generalization ability of data mining method and in order to detect new and unknown attacks. A data mining based IDS uses machine learning and data mining algorithms on a large set of system audit data to build detection models [12]. Hyt *et al.* [12], noted that intrusion detection techniques using data mining have attracted more and more research interests in recent years.

Data mining is the task of discovering interesting patterns from large amounts of data, where the data can be stored in databases, data warehouses, or other information repositories [13]. According to Fayyad *et al.* [6], data mining is the nontrivial process of identifying valid, novel, potentially useful, and ultimately understandable hidden patterns in data. Data mining has two major tasks; predictive modeling and descriptive modeling. As an important application area of data mining, data mining meliorates the great burden of analyzing huge volumes of audit data and realizing performance optimization of detection. Many data mining systems are great in deriving useful statistics and patterns from huge amounts of data, but they are not very smart in interpreting these results, which is crucial for turning them into interesting, understandable and actionable knowledge [14].

Data mining methods such as clustering, classification, and association rule mining are often used to obtain valuable information of network intrusion through analyzing the network data. Clustering can be used in both misuse detection and anomaly detection while classification is mainly used in anomaly detection and is a supervised learning method [15]. An IDS based on classification can classify all the network traffic into either malicious or normal. Association rule mining searches for frequently occurring items from a large dataset and identify correlation relationships or association rules between data items of the large dataset [16].

A Knowledge Based System (KBS) is a sub-field of Artificial Intelligence (AI) that works on knowledge base for effective decision making by imitating the behavior of human expert within a well-defined, narrow domain of knowledge [17].

AI intends understanding of human intelligence and building of computer programs that are capable of simulating or acting one or more of intelligent behaviors. Intelligent behaviors include cognitive skills like thinking, problem solving, learning, understanding, emotions, consciousness, intuition and creativity, language capacity, etc. These days some of the behaviors such as problem solving, learning and understanding are handled by computer programs [11] [7] [18] [19].

Computer systems that try to solve problems in a human expert of the area by using knowledge about the application domain and problem solving techniques are known as Knowledge based system [11] [7].

As stated by Khandelwal [11], the goal of the design of the knowledge-based system is to encapsulate tacit and explicit knowledge in a particular area and code this in a computer in a way that the knowledge of the expert is accessible to novice users.

KBS has come across a variety of approaches based on the knowledge representation methods and the reasoning strategies applied during implementation. Rule based reasoning (RBR) and case based reasoning (CBR) are two popular approaches used in knowledge based systems [20]. According to Prentzas [20], the main advantages of RBR are compact representation of general knowledge, naturalness of representation, modularity and provision of explanations. However, RBR has some drawback such as knowledge acquisition bottleneck, brittleness

of rules, inference efficiency problems, difficulty in maintenance of large rule bases, problem solving experience is not exploited and interpretation problems. On the other hand, Case Based Reasoning has an inability to express general knowledge. Cases, by nature, express specialized knowledge. So, they cannot express general knowledge. This is a disadvantage compared to rule-based systems [20].

As noted in [20], the combination of (two or more) different problem solving and knowledge representation methods is a very active research area in Artificial Intelligence. The aim is to create combined formalisms that benefit from the strength of each method. It is generally, believed that complex problems are easier to solve with hybrid or integrated approaches [20]. One of the popular types of combination involves Rule-Based Reasoning (RBR) and case-based Reasoning (CBR).

The motivation for this research is, as stated by Prentzas [20], the efforts to combine symbolic rules and cases have yielded advanced knowledge representation formalisms. The effectiveness of those approaches stems from the fact that rules and cases are complementary in representing application domains and solving problems. Rules represent general knowledge of the domain, whereas cases represent specific knowledge. Rule-based systems solve problems from scratch, while case-based systems use pre-stored situations to deal with similar new instances.

Further, Azeb [21] attempted to combine rule based reasoning with an existing case based reasoning for oil Drilage Company. The finding of Azeb stated that the combination of rule based reasoning with case based reasoning improve the accuracy of the decisions of the company.

Therefore, the combination of both approaches turns out to be natural and useful. The complementarities of the advantages and disadvantages of the two knowledge representation and reasoning approaches intensify the usefulness of their combination. Using the complementarities nature of the two reasoning systems for knowledge based network intrusion detection is the main motivation for this research.

II. Related Works

The researcher reviews different works done to apply data mining and knowledge based system individually and in combination for network intrusion detection.

Tigabu [22] designed a semi-supervised intrusion detection system using data mining. The dataset for the experiment has been taken from Massachusetts Institute of Technology Lincoln laboratory. After taking the data, it has been preprocessed for filling missed values, removing outliers and resolving inconsistencies Integration of data that contains both labeled and unlabeled datasets, dimensionality reduction, size reduction and data transformation activity like discretization tasks were also done in the study. The system has a prediction accuracy of 96.11% on the training datasets and 93.2% on the test dataset to classify the new instances as normal, DOS, U2R, R2L and probe classes. Data mining has been used for intrusion detection systems due to the fact that they are generally more precise and require far less manual processing and input from human experts. But researches which employed data mining for intrusion detection merely generate patterns and they lack the use of discovered knowledge. Finally, the researcher recommended designing knowledge base system which will add adaptability and extensibility features for intrusion detection.

Abdulkerim [13] attempted to integrate data mining with knowledge based system to design a rule-based intrusion detection system. The aim of the study is utilizing hidden knowledge extracted by employing rule induction algorithm of data mining, specifically JRip from sampled KDDcup‘99 intrusion data set. The integrator application then links the model created by JRip classifier to knowledge based system so as to add knowledge automatically. The system has scored 80.5% overall performance and the researcher recommends further investigations on different algorithms of the same case and the combination/integration of different reasoning systems to refine the knowledge base and boost the advantages of integrating data mining induced knowledge with knowledge based system.

Dawit [23] conduct a research on an integration of data mining with case based reasoning system for network intrusion detection. The system is aiming at utilizing hidden knowledge extracted by employing descriptive algorithm of data mining, specifically K-means clustering from sampled KDDcup’99 intrusion dataset. Clustered case with centroid value is mapped to COLIBRI Studio IDE and then the integrator application creates using Eclipse IDE with JDK 6. The system achieves 89.5% accuracy and further studies suggested to be done to improve the retrieval process by applying ontology based retrieval and integration of case based reasoning with rule based reasoning.

Lidong Wang and Randy Jones [24] presents methods and subsequent evaluation criteria for network intrusion detection, stream data characteristics and stream processing systems, feature extraction and data reduction, conventional data mining and machine learning, deep learning, and Big Data analytics in network intrusion detection.

Jaiganesh [25] Investigates the use of machine learning algorithms for network intrusion detection, in this work; a hybrid intrusion detection system that use classification and clustering algorithms has been designed for distinguishing normal and intrusive events. The major contributions of this research work were the proposal of a hybrid architectural framework for effective intrusion detection.

Therefore, this study focuses exploring a way to automatically use a knowledge acquired using descriptive and predictive mining models for network intrusion detection system that combines case-based reasoning and rule-based reasoning.

III. Architecture of the Proposed NID System

As shown in Fig. 1, the system takes intrusion data set to construct predictive model using classification algorithm such as PART and JRip rule induction as well as REPTree and J48 decision trees. Hidden knowledge available in predictive models are rules, that are used for designing rule-based reasoning system.

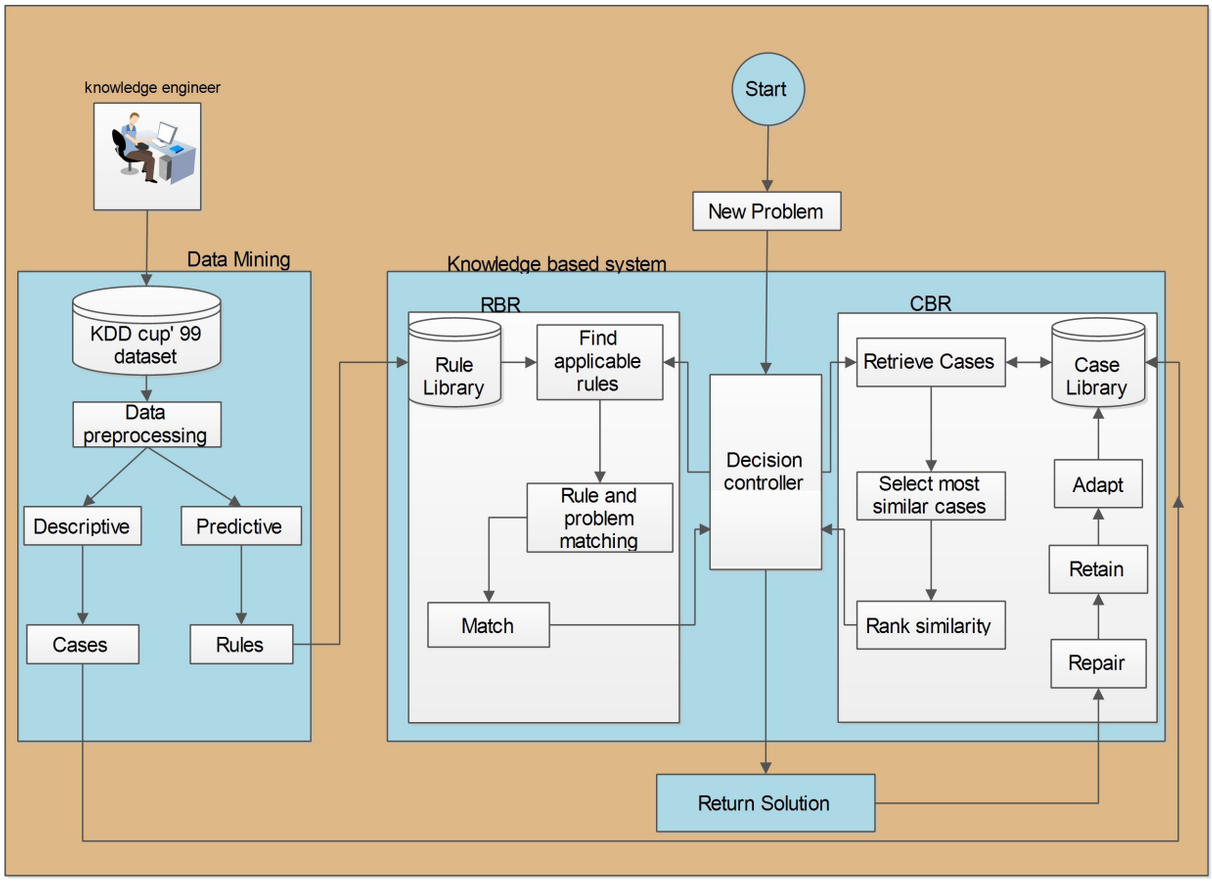

Fig. 1: Network Intrusion Detection System Architecture

In addition, a descriptive model is constructed with the help of clustering algorithms such as K-means, Expectation Maximization (EM) and Make Density Based. From descriptive models, cases are generated for developing case-based reasoning system. Finally, these two systems are combined together for the development of better knowledge based network intrusion detection system.

Among the different conditional combination models (CBR-controller-RBR, RBR-Controller-CBR etc.), the RBR-Controller-CBR model of combination is selected [26] [27].

As new cases or problems happens the combined reasoning system first use RBR to search for the rules/facts that matches with the new query from the rule library. If exact matches found the RBR gives the solution. If not, the query is passed to CBR system.

The CBR system has the capability of retrieve, reuse, revise and retain cases. The retrieval task starts with a new case description and ends when best matching similarity case of attacks or normal cases have been found in the case base. As users enter case description (query) through interface, the system searches the best matching cases from the case base and return possible solution in ranked order to the problems described in the query. If exact matching occurred in between the query and cases in the case base, users can directly derive the solution or if the similarity or matching is partial, then adaptation, revision and reuse tasks are done to make the proposed solution best for the problem at hand. At last, those best modified solution to the problem stored to the case base for future use.

IV. Creating Predictive and Descriptive Model

The KDDcup'99 data set is very huge in size; even after removal of redundant instances the remaining dataset is so huge which requires time and memory space during the mining process. Hence, the sampling is found to be paramount before using data mining to extract patterns from dataset and considerable samples are taken. While sampling, all instances of R2L and U2R are

taken since their size is so small compared to others. The number of instances included in the sample for normal, Probe and DOS is based on their proportion on the cleaned dataset.

The data set was arranged for data mining technique and also suitable for the selected CBR and RBR tools. As a result, the redundant instances in the record were removed during the preparation of dataset for data mining and case features for building the CBR case-base.

The dataset contains four different categories of intrusions such as DOS, R2L, U2R and Probing and also Normal. For this study, 4,898,401 instances were collected from the website and among them 494,427 instances were found non-redundant.

TABLE I: DISTRIBUTION OF DIFFERENT TYPES OF NETWORK INTRUSION INSTANCES

	Total Instances		
Attacks	Before Removing Redundant	After Removing	Percentage
R2L	1,126	1,126	0.22
Probe	41,102	4,107	0.83
U2R	49	49	0.11
DOS	3,883,370	391,458	79.17
Normal	972,780	97,277	19.67
Total	4,898,427	494,427	

Table I presents the distribution of different types of normal and network intrusions instances before and after removing redundant instances.

After data preprocessing and preparation task completed and converted into suitable format for WEKA, the next task is the experimentation involving the selected algorithms. The experiment has two categories, creating predictive model and descriptive model.

The sampled data set contains 54,111 instances containing normal, probe, DOS, U2R and R2L. The data set contains 41 attributes and all of them are involved in all experiments.

A. Creating Descriptive Model

To build a descriptive model and get cases for knowledge representation, three descriptive models involving K-means, Expectation Maximization (EM) and Make Density Based (MD) clustering algorithms are constructed.

TABLE II: PERFORMANCE OF CLUSTERING ALGORITHM

Clustering Algorithm	Correctly Clustered	
	No.	Percentage
K-means cluster	48,396	89.43%
Make Destiny based Cluster	47,419	86.9%
EM cluster	45,300	80.1%

As shown in Table II, to create descriptive models using data mining, the K-means has best performance among the three clustering algorithm and it correctly clustered 89.43% instances. So, the k-means clustering algorithm is used to build cases for the system.

The k-mean cluster similarity is measured in regard to the mean value of the objects in a cluster, which can be viewed as the cluster's centroid or center of gravity. The k-means algorithm takes the input parameter, *k*, and partitions a set of n objects into *k* clusters so that the resulting intracluster similarity is high but the intercluster similarity is low.

B. Creating Predictive Model

To create a predictive model and generate rules using data mining, four predictive models involving J48, REPTree, PART and JRip classifier algorithms are used. J48 and REPTree are decision tree based classifiers in WEKA whereas PART and JRip are rule induction based classifiers.

Rules acquired from the best performing classifier algorithm are used for constructing rule based knowledge base.

So as to develop effective knowledge base system, acquiring relevant rules is paramount. Hence, as shown in Table III below, from the four algorithms *JRip* classifier registered the highest result with 99.4% accuracy in classifying the data set.

TABLE III: PERFORMANCE OF CLASSIFIER

Classifier	Correctly Classified Instances	
	No.	Percentage
PART	54,029	99.848%
JRip	54,078	99.939%
REPTree	53,929	99.654%
J48	54,041	99.870%

As shown below, JRip algorithm generates rules in the form of *(condition)…(conclusion)* format. The *condition* part checks the fulfillments of the constrained values of attributes to conclude the attack type *class='Attack_type'*, for example *class=R2L, class=probe.*

Rule	Condition	Conclusion
Rule 1	root_shell(root_shell>=1) ,duration (duration>=25) .	U2R
Rule 2	num_failed_login (num_failed_logins>=1) .	R2L
Rule 3	(count >= 49) and (dst_bytes <= 0)	DoS
Rule 4	Duration=0)	Normal

Fig. 2: Sample Rules Generated Using JRip from Sample Dataset

Hence, for a certain network incident to be classified, for example as an R2L attack both the antecedent of the rule *((is_*

guest_login = 1) and (duration <= 1)) should be true. That means, if the guest has logged in and stayed less than or equal to one second then that network incident is an R2L attack. If either of them is false then rules related to other attacks are checked.

V. Hybrid Reasoning System

The objective of this study is to use the discovered hidden knowledge using predictive and descriptive data mining techniques to design a combined reasoning system for knowledge based intrusion detection. Predictive model is used for constructing rule-based system and descriptive model is also employed for designing case-based system. These systems are combined systematically to enhance the performance of knowledge-based system for network intrusion detection.

A. Using Descriptive Model for Case Based Reasoning

To integrate automatically the K-means clustering model (constructed by data mining) with Case Based system [23], COLIBRI Studio, Eclipse IDE Version: 3.7.0 with JDK 6.0 is used.

Before applying the first process of integration, the researcher create cbr java packages and then create class for CBR Application and link to k-means clustered result with CBR. After CBR Application and clustered result are applied, the next step could be experiment on COLIBRI Studio.

A CBR application in jCOLIBRI has the capability of retrieve, reuse, revise and retain cases. The retrieval task starts with a new case description and ends when best matching similarity case of attacks or normal cases has been found in the case base. As users enter case description (query) through interface, the system searches the best matching cases from the case base and return possible solutions in ranked order to the problems described in the query. If exact matching occurred in between the query and cases in the case base, users can directly derive the solution or if the similarity or matching is partial, then adaptation, revision and reuse tasks are done to make the proposed solution best for the problem at hand. At last, the best-modified solution to the problem is stored to the case base for future use.

B. Mapping Predictive Model for Rule Based Reasoning

In this study, a predictive model created by JRip rule induction classifier algorithm is used. The algorithm generated 22 rules for the attacks namely; probe, DOS, R2L and U2R attacks and. It can be deduced from the rules that if a certain incident satisfies one of the 22 rules it is an attack, otherwise, it is a normal network incident.

However, to integrate rules extracted by JRip rule induction into knowledge base using prolog programming language, there is a need to convert the structure from (condition)=> (conclusion) format (see Fig. 2) to (conclusion):- (condition) as shown below.

Rule 1: attack (u2r):-

root_shell(root_shell>=1), duration (duration > = 25).

Rule 2: attack (r2l):

num_failed_logins(num_failed_logins>=1).

Rule 3: attack (DoS):-

(count >= 49) and (dst_bytes <= 0) => class=DOS (6977.0/0.0).

Rule 4: Normal:- (Duration=0) => class=normal(21999.0/4.0)

Therefore, the above rule has been formatted as: *attack (u2r):- (is_guest_login = 1),(duration =< 1).*

Once rule-based and case-based reasoning's are constructed from predictive models and descriptive models respectively, an attempt is made to combine during reasoning.

C. Combination of Case Based with Rule Based Reasoning

During combining case-based and rule-based reasoning, an experiment is made to decide the order of using them for network intrusion detection.

Basically, there are two ways of combining the two reasoning techniques i.e. RBR-Controller-CBR and CBR-Controller-RBR [26]. While the first one applies, RBR to find a solution for the query, otherwise CBR is used to find similar cases with the query. The reverse is true for the case CBR-Controller-RBR.

To choose which model of combination is more preferable in the network detection process, an experiment is conducted using the same test dataset and the performance of individual reasoning rule-based reasoning and case-based reasoning as well as their combined reasoning (CBR-controller-RBR and RBR-controller-CBR) are compared.

Table IV below depicts the comparison of the performance of the reasoning mechanisms. The comparison is made in terms of the percentage of correctly and incorrectly identified type of network attacks.

Table IV: Experimental Result of Individual and Combined Reasoning Methods

Reasoning Method	% of Correctly Identified	% of Incorrectly Identified
RBR	81.2%	18.8%
CBR	86.3%	13.7%
CBR-RBR	86.3%	13.7%
RBR-CBR	87.66%	12.34%

From the experimental result, a combination reasoning model of RBR-controller-CBR (RBR-CBR) performs better than the other model of reasoning by achieving 87.66% correctly classified network intruders. This is because, as stated in [27]

[26] the following are the advantages of RBR-Controller-CBR model of combination.

- RBR module uses more general knowledge to solve situations not reflected in CBR. Therefore, to solve problems of a given area, it is common to start from the general knowledge.
- RBR is a collection of rules/facts and used exact matching of problems to solve. If the problem matches, the solution will be 100% acceptable or correct, but CBR measures similarity to identify cases near to the given query.
- RBR allows shortcuts in reasoning. If appropriate rules can be found, problems can often be solved in much less time than it would take to generate a solution from cases.

So, in this study, the RBR-CBR combination approach is selected and used for implementing a prototype network intrusion detection system.

VI. Network Intrusion Detection Process

The intrusion detection process is undertaken by interacting with the user through presenting a serious of questions for the user. The system asks the user a series of yes/no questions (see Fig. 3), based on which it applies RBR first to detect network incident as Probe, DOS, U2R and R2L attacks, it displays the result. Otherwise, the query directly moves to the CBR for recommending the most similar case with the query.

Fig. 3 shows, the user interface that asks experts to respond 'Yes or 'No' contains some of the rules generated by the JRip algorithm to determine the type of attack U2R.

On the above interface, It allows the user to select the radio button as per the new query and hit the check button, if the condition is satisfied, the type of attack is displayed at the provided space, like " Attack type: U2R".

(root_shell>=1)and (duration>=25)=>class=U2R

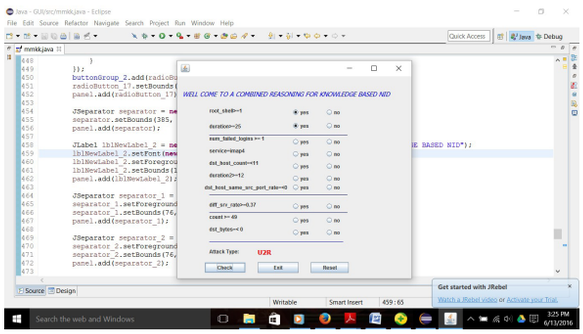

Fig. 3: A Combined User Interface Detecting Intruder Type U2R

If one of the selection of the attribute is "no", the RBR is not able to identify the type of attack. Hence, the message "The RBR cannot identify the type of attack, please consult the CBR" is displayed. If "ok" is selected it automatically goes to the CBR query dialog box. The CBR query interface is displayed automatically and diagnoses the type of attack.

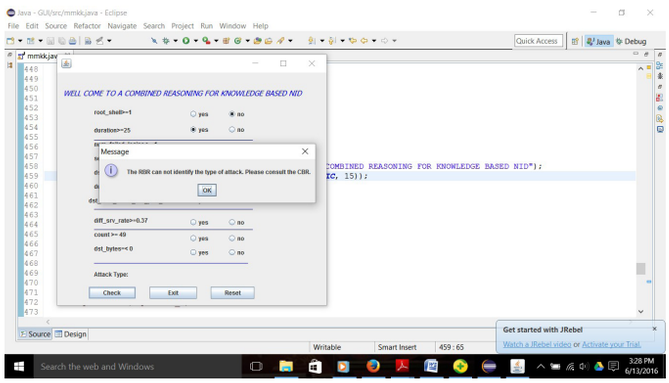

Fig. 4: User Interface of Combined System Showing the RBR Cannot Identify the Type of Attack

The CBR, consider the value of the attributes measure to similarity, retrieve cases and display the type of attack. Retain and adapt is also took place after the necessary action is taken by the network administrator.

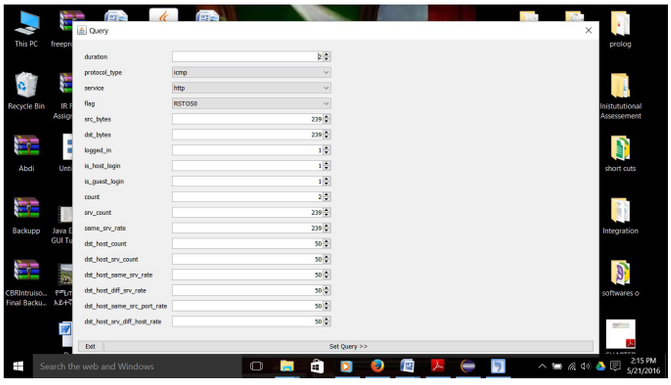

Fig. 5: The Query Interface of the CBR

VII. Evaluation of the System

In order to assure the combined NID meets the requirement it is developed for system performance testing is performed.

Confusion matrix is used for comparing the performance of Combined-NID based on domain expert's judgment (See Table V).

TABLE V: CONFUSION MATRIX FOR EVALUATION OF COMBINED-NID COMPARED TO EXPERTS' FEEDBACK

	Performance of Combined Reasoning System						
Domain experts feed-back	Attack Class	Normal	Probe	DOS	R2L	U2R	Total
	Normal	6	0	0	0	3	9
	Probe	0	5	0	0	0	5
	DOS	0	1	5	0	0	6
	R2L	0	0	0	6	0	6
	U2R	0	0	0	0	4	4
	Total	6	6	5	6	7	30

As shown in Table V, the system has correctly detected 28 test instances out of 30 to their correct class. This means the system has 93.33% detection accuracy. With the exception of 'Normal' and 'DOS', there is a 100% correct classification. Most of the mis-classification happens with 'Normal' class and 'U2R'. As domain expert suggests, this misclassification happens because of their behavioral similarities.

TABLE VI: PERFORMANCE EVALUATION BASED ON PRECISION, RECALL AND F-MEASURE

Attack Class	Precision	Recall	F-Measure
Normal	0.89	0.87	0.80
Probe	0.83	1	0.90
DOS	1	0.87	0.93
R2L	0.8	0.85	0.66
U2R	1	1	1
Average	0.90	0.91	0.905

Precision and Recall are also employed to evaluate the system performance apart from detection accuracy. Recall is the proportion of real positive cases that are correctly predicted positive. Precision denotes the proportion of predicted positive cases which are correctly real positives [28].

The system has registered its highest 100% Recall for Probe and U2R attacks and 87% for Normal and DOS, and 85% for R2L attacks. The system registered its highest 100% Precision for DOS and U2R respectively. The system has 89% Precision for Normal, 83% for Probe and 80% for R2L. In general, the system has performed well in both Recall (91%) and Precision (90%), for identifying network incidents accurately with F-measure 90.5%.

VIII. CONCLUSION

In this paper, a combined reasoning system for knowledge-based network intrusion detection has been designed. The combination of two reasoning systems, i.e. Case Based Reasoning (CBR) and Rule Based Reasoning (RBR) system is employed. We use descriptive modelling to construct cases for a case based reasoning a descriptive modelling and for generating rules for a rule based reasoning predictive modelling are used. The KDD CUP 1999 Dataset has been used to train and test the combined reasoning system for knowledge-based-NID proposed in this research work.

The performance of the system shows that it achieves 90.5% accuracy with an average Precision and Recall of 90% and 91% respectively. User acceptance testing also shows on the average an excellent acceptance. This shows the system has registered a promising result.

The strength of this study is the reasoning system to detect intruders in a combined manner so that the probability of the intruders to be detected is high. In combination of RBR and CBR, due to their interchangeable nature, they provides effective knowledge representation, problem solving power, and exceeding one's weakness with the other. The fruit of the complementary nature of the reasoning systems also proposed here for network intrusion detection. Further, we are working to develop an intelligent system that prevents network intruders.

REFERENCES

[1] S. K. Miller, "An introduction to computer security," *IEEE Computer*, vol. 38, no. 34, 2001.

[2] A. H. Fares, and M. I. Sharawy, "Intrusion detection: Supervised machine learning," *Computing Science and Engineering*, vol. 5, no. 4, pp. 305-313, December 2011.

[3] A. Usman, H. Sajjad, N. Salman, and U. Obaid, "A survey of intrusion detection and prevention techniques," *International Conference on Information Communication and Management*, vol. 16, pp. 23-36, 2011.

[4] M. Sumit, M. Mary, J. Anupam, and F. Tim, "A knowledge-based approach to intrusion detection modeling," in *Computer Science and Electrical Engineering*, Maryland, pp. 19-30, 2006.

[5] D. Hervé, "An introduction to intrusion-detection systems," in *Proceedings of Connect'2000,* Zurich, pp. 1-18, IBM Research, IBM Reseach Laboratory, 2002.

[6] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, "Knowledge discovery and data mining: Towards a unifying framework," in *Knowledge Discovery and Data Mining*, California, pp. 82-88, 1996.

[7] D. Hassen, "Integrating descriptive modelling with case based reasoning in network intrusion detection," Machine Learning, School of Information Science: Addis Ababa University, M.Sc. Thesis, 2015.

[8] Jaiganesh "Investigation on machine learning algorithms for network intrusion detection system," Department of Computer Science & Engineering, Manonmaniam Sundaranar University, Ph.D. Dissertation, July 2014.

[9] P. Dokas, L. Ertoz, V. Kumar, A. Lazarevic, A. J. Srivastava, and P. N. Tan, "Data mining for network intrusion detection," IEEE, 1987.

[10] *Intrusion Detection Systems*, 6th ed., New York: Information Assurance Tools Report IATR, 2009.

[11] Khandelwal, "Knowledge based systems, problem solving competency and learnability," Suresh Gyan Vihar University, Department of Computer Science, 2014.

[12] H. A. Nguyen, and D. Choi, "Application of data mining to network intrusion detection: Classifier selection model," in *APNOMS*, Berlin, pp. 399-408, 2008.

[13] M. Abdilkerim, "Towards integrating data mining with knowledge based system: The case of network intrusion detection," School of Information Science, Addis Ababa University, Addis Ababa, M.Sc. Thesis, 2013.

[14] P. Carsten, "Integrating and updating domain knowledge with data mining," Leipzig Graduate School of Management, vol. 3, pp. 50-62, June 2000.

[15] J. Schäfer, "Capture, distribution and evolution of information needs in a process-oriented knowledge management environment," University of Kaiserslautern, Department of Computer Science, 2003.

[16] M. K. Patond, and Deshmukh, "Survey on data mining techniques for intrusion detection system," *International Journal of Research Studies in Science, Engineering and Technology*, vol. 1, no. 1, pp. 93-97, 2014.

[17] R. Kumar, *Research Methodology - A Step-by-Step Guide for Beginners*, 2nd ed., Singapore: Sage, 2005.

[18] R. S. Engelmore, and E. Feigenbaum, "Japanese technology evaluation center, knowledge based systems in Japan," 21 March 2009. Available: http://www.wtec.org/loyola/kb/toc.htm

[19] A. J. Champandard, "AI depot, Artificial Intelligence," 23 October 2008. Available: http://aidepot.com/Intro.html

[20] J. Prentzas, and I. Hatzilygeroudis, "Categorizing approaches combining rule-based and case-based reasoning," *Expert Systems*, vol. 24, no. 2, pp. 97-122, May 2007.

[21] A. Bekele, "Integrated case based and rule based reasoning for decision support," Norwegian University of Science and Technology, Department of Computer and Information Science, M.Sc. Thesis, 2009.

[22] T. Dagne, "Constructing predictive model for network intrusion detection," School of Information Science, Addis Ababa University, Addis Ababa, M.Sc. Thesis, 2012.

[23] D. Hassen, "Integrating descriptive modelling with case based reasoning in network intrusion detection," School of Information Science, Addis Ababa University, M.Sc. Thesis, 2014.

[24] L. Wang, and R. Jones, "Big data analytics for network intrusion detection: A survey," *International Journal of Networks and Communications*, vol. 7, no. 1, pp. 24-31, 2017. DOI: 10.5923/j.ijnc.20170701.03

[25] Jaiganesh, "Investigation on machine learning algorithms for network intrusion detection system," Department of Computer Science & Engineering, Manonmaniam Sundaranar University, Ph.D. Dissertation, July 2014.

[26] K. Khandelwal, and D. P. Sharma, "Hybrid reasoning model for strengthening the problem solving capability of expert systems," *International Journal of Advanced Computer Science and Applications (IJACSA),* vol. 4, no. 10, pp. 88-94, 2013.

[27] E. L. Rissland, and D. B. Skalak, "Combining case-based and rule-based reasoning: A heuristic approach," in *Eleventh International Joint Conference on Artificial Intelligence (IJCAI'89),* pp. 524-530, 1989.

[28] D. M. W. Powers, "Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation," *Journal of Machine Learning Technologies*, vol. 2, no. 1, pp. 37-63, 2011.