Fuzzy Rough Set based Feature Selection and Enhanced KNN Classifier for Intrusion Detection

استخدام نظرية المجاميع العشوائية المضببة القائمة على ميزة الاختيار وخوارزمية تصنيف الجار الأقرب المحسنة لكشف التسلل في قواعد بيانات شبكات الانترنيت

Osamah Mohammed Fadhil
Computer System Department, Qurna Technical Institute, Southern Technical University, Basra, Iraq
Osama.alyasiri@gmail.com

Abstract:

Intrusion detection systems are used to detect and prevent the attacks in networks and databases. However, the increase in the dimension of the network dataset has become a major problem nowadays. Feature selection is used to reduce the dimension of the attributes present in those huge data sets. Classical Feature selection algorithms are based on Rough set theory, neighborhood rough set theory and fuzzy sets. Rough Set Attribute Reduction Algorithm is one of the major theories used for successfully reducing the attributes by removing redundancies. In this algorithm, significant features are selected data are extracted. In this paper, a new feature selection algorithm is proposed using the Maximum dependence Maximum Significance algorithm. This algorithm is used for selecting the minimal number of attributes has been from KDD data set. Moreover, a new K-Nearest Neighborhood based algorithm is proposed for classifying data set. This proposed feature selection algorithm considerably reduces the unwanted attributes or features and the classification algorithm finds the type of intrusion effectively. This system is very efficient in detecting attacks and effectively reduces the false alarm rate. The proposed feature selection and classification algorithms enhance the performance of the IDS in detecting the attacks.

Keywords: Rough Set, Fuzzy Set, Feature Selection, Classifications and Intrusion Detection.

الملخص

تستخدم انظمة كشف التسلل لكشف ومنع الهجمات في قواعد بيانات الشبكات. ومع ذلك، فان التوسع في مجالات قواعد بيانات الشبكات أصبح يمثل مشكله رئيسية هذه الايام. خاصية اختيار الميزة استخدمت لتقليل من حده هذه المشكلة التي تتمثل في ضخامة البيانات. خوارزمية اختيار البيانات التقليدية استندت على نظرية الأنظمة العشوائية ونظرية الأنظمة العشوائية المجاورة والمجاميع المضببة. خوارزمية تقليل الأنظمة العشوائية هي واحده من النظريات الرئيسية التي نجحت في تقليل السمات من خلال حذف الزوائد في هذه الخوارزمية، فان صفات مهمة تم تجريبها على بيانات مختصرة.

في هذه البحث، تم اقتراح خوارزمية اختيار سمات جديده باستخدام الحد الأقصى اعتمادا على خوارزمية الدلالة القصوى. وتستخدم هذه الخوارزمية لاختيار الحد الأدنى لعدد من صفات مجموعة البيانات *KDD*. وعلاوة على ذلك، تم اقتراح خوارزمية جديدة (*K-Nearest Neighborhood*) لتصنيف مجموعة البيانات. خاصية خوارزمية الاختيار المقترحة قللت وبشكل ملحوظ من البيانات الغير مرغوب فيها، ووجدت خوارزمية التصنيف نوع التسلل بشكل فعال. وهذا النظام يعتبر فعال جدا في كشف الهجمات ويقلل من فعالية معدل انذار كاذب. ان اختيار خوارزمية السمات المقترحة وتصنيفها عزز من أداء نظام كشف التسلل.

1. Introduction

Intrusion detection is needed in today's computing environment because it is impossible to keep pace with the current and potential threats and vulnerabilities in computing systems. The networking environment is constantly evolving and changing due to the advances in web and internet technologies. To make matters worse, threats and vulnerabilities in the environment is also constantly evolving. An Intrusion detection system can be used to assist in managing threats and vulnerabilities in system. Threats occur due to people or groups who have the potential to compromise system. Moreover, the hackers have become a serious threat to many companies in the software field and those in other fields also suffer from this problem. Vulnerabilities are weaknesses in the systems, which are exploited by the hackers to compromise the system. New vulnerabilities are introduced every time when the technology develops and hence brings a new technology, product, or system brings with it a new generation of bugs and unintended conflicts arise. Moreover, the possible impacts from exploiting these vulnerabilities are constantly evolving. An intrusion may cause production downtime, sabotage of critical information, and theft of confidential information, cash, or other assets.

It may even create negative public relations that may affect a company's growth. Intrusion detection products are able to assist in protecting a company from intrusion by expanding the options available to manage the risk from threats and vulnerabilities. Intrusion detection capabilities can help a company secure its information. The system can be used to detect an intruder, identify and stop the intruder, support investigations to find out how the intruder got in, and stop the exploit from use by future intruders. The correction should be applied across the enterprise to all similar platforms. Intrusion detection products can become very powerful in the information security. Different methods of attack are listed below:

Masquerading: It includes using a stolen username/password or sending a TCP packet with a forged source address.

Abuse of Feature: Includes filling up a disk partition with user files or starting hundreds of telnet connections to a host to fill its process table.

Implementation Bug: A bug in a trusted program might allow an attack to proceed.

System Mis-configuration: An attacker can exploit errors in security policy configuration that allows the attacker to operate at a higher level of privilege than intended.

Social Engineering: An attacker may be able to coerce a human operator of a computer system into giving the attacker access. A bug in the implementation of the TCP stack on some systems makes it possible to crash the system by sending it a carefully constructed malformed TCP packet.

In feature selection in IDS also known as variable selection, attribute selection or variable subset selection, is the process of selecting a subset of relevant features for use in model construction. The central assumption when using a feature selection technique is that the data contains many redundant or irrelevant features. Redundant features are those which provide no more information than the currently selected features, and irrelevant features provide no useful information in any context. Feature selection techniques are a subset of the more general field of feature extraction. Feature extraction creates new features from functions of the original features, whereas feature selection returns a subset of the features. Feature selection techniques are often used in domains where there are many features and comparatively few samples.

Feature selection is also useful as part of the data analysis process, as it shows which features are important for prediction, and how these features are related. The choice of evaluation metric heavily influences the algorithm. It is these evaluation metrics which distinguish between the three main categories of feature selection algorithms: wrappers, filters and embedded methods. Other popular approach is the Recursive Feature Elimination algorithm, commonly used with Support Vector Machines to repeatedly construct a model and remove features with low weights.

In this paper, we propose new IDS for enhancing security that utilizes the information on KDD dataset. The contributions of this paper are as follows.

- We define and select important feature from the collected dataset. We also suggest a way of removing redundant attributes.
- We provide how to determine the correct user and set alarm to attackers by using the selected features.
- We propose a new classification algorithm for enhancing the security.

This remainder of this paper is organized as follows. In Section 2, related studies are reviewed, and in Section 3, an overview of the proposed IDS is given. Section 4 covers feature selection technique, and Section 5 discusses the proposed classification algorithm of the experimental results, and in Section 6, the conclusions work are presented.

2. The Related work

In this section, the related work on feature selection, feature classification and Intrusion Detection System are reviewed.

2.1 Feature Selection

There are many works in the literature that discuss about Feature Selection [1][2][3]. Feature selection is an important technique in selecting the best attributes for a given data set. Select the features that do not affect the set when a particular attribute is removed. Therefore, it is necessary to propose an algorithm that efficiently reduces the attributes.

Jinbo Bi et al described a methodology for performing variable selection using Support Vector Machines (SVMs). It constructs a series of linear SVMs to generate linear models [1]. The distribution of the linear model weights provides a mechanism for ranking the effects of variables. It reduces the number of variables and outperforms SVMs using all attributes and also using the attributes selected according to correlation coefficients. The visualization of the final result models is useful for understanding the role of underlying variables. Xiubo Geng et al use the value of the feature to rank the training instances and also define the ranking accuracy in terms of performance measures [3]. They also define the similarities between them as the correlation of the features. Their results show that their method outperforms traditional methods for ranking the feature.

Carla E. Brodley and Jennifer G. D [4] defined an automated feature subset selection algorithm for unlabeled data. It explores the feature selection problem through Feature Subset Selection using Expectation-Maximization Clustering (FSSEM) and uses scatter separability and maximum likelihood performance criteria for evaluating candidate feature subsets. It also presents a normalization schema that can be applied to any criterion. George Forman [5] analyzed the feature selection in text domains to make learning efficient. His paper presents a comparison of twelve methods evaluated on a benchmark gathered from routers. The bi-Nominal separation widened the margin with high class screw. This uses the information gain ratio and provides best results in text feature selection.

Lei Yu and Huan Liu [6] identified that feature relevance alone is sufficient for efficient feature selection. This new frame work decouples the relevance analysis and redundancy analysis. It declares that removing redundancy is also important in the feature selection techniques. The feature selection algorithm proposed by Geetha Ramani [7] deals with the statistical method for analyzing the voluminous KDD cup dataset. Since the KDD dataset is very large, we cannot use the hierarchical cluster analysis. The proposed k-means classification algorithm increases the classification accuracy and reduces the computation time. Wei-Zhi Wu et al [8] presents a framework for the study of fuzzy rough sets in which both constructive and axiomatic approaches are used. In axiomatic approach, various classes of fuzzy rough approximation operators are characterized by different sets of axioms.

Richard Jensen and Qiang Shen [9] explained that the rough set is reliant upon a discredited dataset i.e. important information may be lost as a result of dissertation. They proposed a new dimensionality reduction technique that uses a hybrid variant of rough sets (fuzzy-rough sets) to avoid this information loss. Sindhu et al [10] proposed a wrapper based feature selection algorithm in order to develop an IDS. Their approach is better for selecting features and provides high detection rate. Chun-Wei Tsai et al [24], proposed an incremental particle swarm optimization algorithm to enhance the performance of IDS using feature selection.

2.2 Feature Classification Techniques

Iosif-ViorelOnut and Ali A. Ghorbani [12] presented a feature classification schema for network intrusion detection that intends to provide better understanding of the features extracted from the network packets. Further, a feature extractor is provided that extracts the data sets with respect to the attacks and also highlights some sensitive features to attacks. Sannasi Ganapathy et al [13] provided a survey on intelligent techniques for feature selection and classification for intrusion detection in networks. Their discussion is based on intelligent software agents, neural networks, genetic algorithms, neuro-genetic algorithms, fuzzy techniques, rough sets, and their proposed a particle swarm intelligence algorithm. It identifies and prevents network intrusions in order to provide security to the Internet and to enhance the quality of service. Intelligent rule based algorithms for enhancing the multiclass support vector machine for effective feature selection was also proposed.

Huaguang Zhang et al [14] proposed a new algorithm for pattern classification using a new data core based Fuzzy Min-Max Neural Network. Comparing with Fuzzy Min-Max Neural Network Classifier with Compensatory Neuron, this work is different since the hyper box can be expanded and can overlap repeated with the previous hyper boxes. Therefore, it generates minimum number of hyper boxes for rule extraction.

2.3 Works on Intrusion Detection System

Many works are present in the literature about IDS [15][16][17]. Debar et al [15] developed a Neural Network (NN) model for IDS. The advantage of this proposed system is that the deviation from the normal behavior of the user could be easily diagnosed fairly and quickly by the NN. Ahmed Patel et al [16] have implemented the Intrusion detection and Prevention System. The distributed and open technology of cloud services as in mandatory target for potential cyber-attacks by intruders. This proposed work explores a prevention technique in intrusion detection system to detect and prevent the intrusions in cloud computing systems.

Moradi and Zulkernine [17] presented a new IDS that uses Artificial Neural Network (ANN) for effective intrusion detection. They have taken care in adding new agents in such a way that the failure of one agent does not degrade the overall detection performance of the network. When compared with a centralized system, the distributed framework proposed by them is cost effective and efficient. Sarasamma et al [18] proposed a novel multilevel hierarchical Kohonen networks to detect intrusions in networks. In their work, they randomly selected data points from KDD Cup 99 to train and test the classifier. The results obtained by them show that the hierarchical networks in which each layer operates on a small subset of the feature space is superior to a Kohenen network operating on the entire feature space in detecting various kinds of attacks.

Jianping Li et al [19] proposed a new method based on Continuous Random Function (CRF) for selecting appropriate feature sets to perform network intrusion detection. Moreover, it uses network connection information data sequence and the feature sets to detect attack and also for the discovery of abnormal behaviour. The main advantage of their model is that it is practical reliable and efficient.

2.3 Works on Classification

There are many classification algorithms based on SVM that are found in the literature for IDS. For example, an algorithm called Tree Structured Multiclass SVM has been proposed by Snehal A.Mulay et al [20] for classifying data effectively. Their paper proposed a decision tree based algorithm to construct a multiclass IDS which is used to improve the training time, testing time and accuracy of IDS. The performance of unsupervised anomaly detection approaches achieve higher detection rate over supervised approach. Also, unsupervised approach have high false positive rate over supervised approach. However, using unsupervised anomaly detection techniques the system can be trained with unlabeled data and can be made capable of detecting previously unseen attacks.

Manindharsingh et al [21] implemented an Intrusion Detection System to prevent the attacks in mobile Adhoc networks. In Adhoc on Demand distance Vector protocol insider attacks are very common. Hence the proposed work is developed to deduct and isolate these attacks. This work provides the stable and effective observations of attacks which is applicable in the real environment for mobile Adhoc devices. Mohammad Sazzadul Hoque et al [22] presented an intrusion detection system by applying Genetic Algorithm (GA) to efficiently detect the various types of intrusions. To improve the security in communication over the internet or other networks, many Intrusion Detection mechanisms has been implemented but unfortunately any of the system is not completely flawless. So their proposed work combines the intrusion detection system with genetic algorithm to efficiently detect the intrusions in network.

Kartit et al [23] concentrated on the research area in the field of information security. This has been improved to network security today. Security mechanisms are to find the unauthorized users as well as the authorized users. John M. Fossaceca et al [25], proposed a novel algorithm for Multiple Adaptive Learning which combines Multiple Kernel Boosting with the Multiple Classification in the Reduced Kernel. Using this approach, the authors to improve the detection rate and reduced the false alarm rate in intrusion detection system. Riyanat Shittua et al [26], proposed a new framework for Analyzing Intrusion Alerts using classification. They carried out the experiments with industrial partners of the British Telecom Security Practice and showed a false positive rate 97%. Adel Sabry Eesaa et al [27], proposed a new feature selection algorithm based on subsets and used it in decision tree classifier. Nader P et al, the support vector data description and kernel principle component analysis for detecting several types of cyber-attacks using classification.

Abduvaliyev A et al [28], proposed an Intrusion Detection Systems (IDS) for wireless sensor networks. They focused on both anomaly and misuse detection in their classification model. Yunho Lee et al [29] [29], proposed an improved reputation-based intrusion detection system for securing ad-hoc networks based on classification. They analyzed its performance in a real environment through implementation. Thongkanchorn K et al [30], investigated the performance and the detection accuracy of three popular open-source intrusion detection systems namely Snort, Suricata and Bro. Their experiments showed Bro provides better performance than other IDS systems due to the use of better classifier.

Comparing with all the works present in the literature the IDA proposed in this paper different in many ways. First, most of the existing systems are developed for securing the networks. The existing systems such as a SNORT are network based IDSs, on the other hand, the proposed system is a host based IDS made for systems. However, the proposed systems uses KDD data set for effective evaluation. Third, the existing systems such as Suricata and Bro provide accuracy in the range of 82% to 90%. However, the proposed system provides more than 98% of detection accuracy and hence, reduces the false positive rate. Finally, the main advantage of the proposed system is its accuracy in feature selection and classification.

3. Proposed System Architecture

The architecture of the system proposed in this paper consists of six major components namely data set, Data Collector and Preprocessing module, Classification, Mapping module and enhanced KNN classifier as shown in the figure 1.

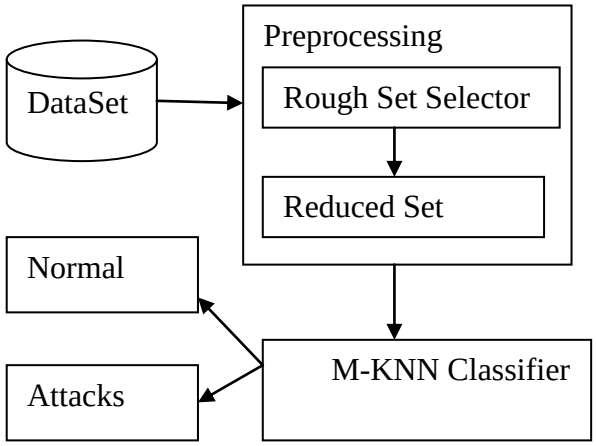

Figure 1, Intrusion Detection System

In this work, the KDD data set is used to carry out the experiments. This is a standard set of data which was audited, and includes a wide variety of intrusions simulated in this network environment, was provided. A connection is a sequence of TCP packets starting and ending at some well-defined times, between which data flows to and from a source IP address to a target IP address under some well-defined protocol. Each connection is labelled as either normal, or as an attack, with exactly one specific attack type. Each connection record consists of about 100 bytes.
Attacks fall into four main categories:

DOS: denial-of-service, e.g. synchronous flood.

R2L: unauthorized access from a remote machine, e.g. guessing password.

U2R: unauthorized access to local super user (root) privileges, e.g., various "buffer overflow" attacks.

Probing: surveillance and other probing, e.g., port scanning.

It is important to note that the test data is not from the same probability distribution as the training data, and it includes specific attack types not in the training data. This makes the task more realistic. Some intrusion experts believe that most novel attacks are variants of known attacks and the "signature" of known attacks can be sufficient to catch novel variants. The dataset contains a total of 24 training attack types, with an additional 14 types in the test data only. **Table1**.shows the different types of attack are present in the KDD dataset.

Table 1. Different types of Attacks in KDD Dataset

Attack Classes	Attacks
Probing	ipsweep, nmap, portsweep, satan
Denial of Service (DOS)	back, land, neptune, pod, smurf, teardrop
User to Root (U2R)	buffer_overflow, perl, loadmodule, rootkit
Remote to User (R2L)	ftp_write,guess_passwd,imap,multihop,phf,spy, warezclient, warezmaster

Feature selection is a term commonly used in data mining to describe the tools and techniques available for reducing inputs to a manageable size for processing and analysis. Feature selection implies not only cardinality reduction, which means imposing an arbitrary or predefined cutoff on the number of attributes that can be considered when building a model, but also the choice of attributes, meaning that either the analyst or the modeling tool actively selects or discards attributes based on their usefulness for analysis.

The ability to apply feature selection is critical for effective analysis, because datasets frequently contain far more information than is needed to build the model. For example, a dataset might contain 500 columns that describe the characteristics of customers, but if the data in some of the columns is very sparse you would gain very little benefit from adding them to the model. If the unneeded columnse kept while building the model, more CPU and memory are required during the training process, and more storage space is required for the completed model.

Even if resources are not an issue, it is necessary to remove unneeded columns because they might degrade the quality of discovered patterns, for the following reasons.

- Some columns are noisy or redundant. This noise makes it more difficult to discover meaningful patterns from the data.
- To discover quality patterns, most data mining algorithms require much larger training data set on high-dimensional data set. But the training data is very small in some data mining applications.

If only 50 of the 500 columns in the data source have information that is useful in building a model, they can be left out of the model, or it is better to use feature selection techniques to automatically discover the best features and to exclude values that are statistically insignificant. Feature selection helps solve the twin problems of having too much data that is of little value, or having too little data that is of high value.

In this proposed work, dimensionality is reduced using a MDMS feature selection algorithm and the attacks are detected under enhanced KNN classification.

- The proposed is the dimensionality reduction in fuzzy rough sets using Minimum Dependence Maximum Significance (MDMS) algorithm.
- MDMS algorithm calculates the dependency value and rank the goodness of the feature.
- Significant features are only considered and thereby attributes are reduced.

Feature Selection Algorithm is to select the best features from a huge data set so as to reduce the computational time. Feature selection is to choose a subset of input variables by eliminating features, which are irrelevant or of no predictive information. Among the huge number of attributes or features present in real-life data sets, only a small fraction of them are effective to represent the data set accurately. Feature selection techniques are to discover the best features automatically and to exclude values that are statistically insignificant.

Our proposed work reduces the computational time for detecting the attacks. MDMS algorithm reduces the number of attributes by selecting only the significant attributes. Significance is ranked on the basis of dependency value. Maximum of dependency higher the significance value and those attributes are added to the significant features. The above steps are repeated until we the attributes are reduced to the significant values.

The performance analysis of this proposed work is done using the enhanced KNN Classifier. Enhanced KNN compares the training data set and testing data set and finally evaluates the performance of our reduced attributes with the real world data set. With the enhanced KNN classifier the computational time is reduced and thereby overlapping of classes can be avoided. Enhanced KNN compares the training data set and testing data set and finally evaluates the performance of our reduced attributes with the real world data set.

4. Extraction Feature Selection Algorithms

4.1 Feature Selection

This section explains the concept of feature selection, feature classification. **Fig 2.** describes the algorithm used for feature selection.

Step 1: Calculate the equivalence classes for each conditional attributes and also for the decision attributes by eq.1.

$$IND(P) = \{(x_i, x_j) \in U \times U \; \forall \; a \in P, f(x_i, a) = f(x_j, a) \qquad (1)$$

IND- indiscernibility, U – universal set, x_i ,x_j are the equivalence classes and P- subset in a data set.
Indiscernibility relation is to calculate the equivalence classes. The equivalence classes are the classes that have the similar characteristic objects. Objects of the same rank can be grouped into a single class. Likewise the objects are classified under various equivalence classes.

Step 2: Remove the redundant attributes. Data set may contain useful as well as unwanted information that has no predictive measure. Removing those redundant or unwanted attributes may help us in reducing the computational time.

Step 3: Using eq.2, 3 calculate the fuzzy Approximations to calculate the positive region of the attribute.

$$\mu\overline{P}(F_i) = \sup\{\mu F_i(x), \quad \mu\overline{P}X(F_i)\} \qquad (2)$$

$$\mu\underline{P}(F_i) = \sup\{\mu F_i(x), \quad \mu\underline{P}X(F_i)\} \qquad (3)$$

μ - membership function, F_i – Fuzzy equivalence class, sup – supremum
$\underline{P}$ - Fuzzy lower approximation,$\overline{P}$ - Fuzzy upper approximation.
Fuzzy lower and upper approximations are to calculate the positive region of the conditional attribute with respect to the decision attribute and then the dependency.

Step 4: Calculate the positive region of the conditional attribute for decision attribute.
$POS = U\underline{C}X_i$, X_i is the *i*th equivalence class.
POS – Positive region of the attribute
The positive region of the conditional attributes can be computed from the fuzzy lower approximation.

Step 5: Calculate the dependency value with the calculated positive region of the attribute.

$$\gamma_c = \frac{|POS_c(D)|}{|U|} \qquad (4)$$

γ_c– Dependency value of the attribute.
Dependency value is calculated to find the significance of the attributes. It shows the importance of the feature on predicting the result.

Step 6: Categorize the dependent attributes of high value into the significant, insignificant and dispensable set.
Significant set has the attributes of high dependent value. These are the most important attributes in predicting the result.

$$S_i = \{A_j \mid \sigma(A_j, D) > \delta_i\} \qquad (5)$$

Insignificant set has the attributes that has zero level predictive measure. So this set is rejected or omitted.

$$I_i = \{A_j \mid \sigma(A_j, D) < -\delta_i\} \qquad (6)$$

Dispensable set has the attributes that has either predictive or non-predictive measure. This has to be run into loop for finding the dependent value and considered to a significant extent.

$$D_i = \{A_j \mid \sigma(A_j, D) \leq -\delta_i\} \qquad (7)$$

Finally the attributes are reduced to the significant set.
A_j – Conditional Attribute value, D – Decisions Attribute value, σ – Significance value and δ – Threshold value.

Input: F (f1, f2, …fn) Selected Features
Output: Return Labelled Classes

1. Set k, the number of neighbors
2. Calculating the Nearest Neighbors from F
3. Initialize E to NULL
4. For i=1 to n do begin
5. Calculate the Euclidean distance from node to member x_i
6. Compute the Fuzzy membership value
7. Apply fuzzy rule to check nearness.
8. If i < k then
9. Add member to set E
10. Else if i=k then
11. Member closer to node then assign member as the nearest neighbors.
12. Else
13. Delete the member from the set.
14. End for
15. Read c;
16. For i=1 to c
17. Calculate the membership $\boldsymbol{\mu_i}$
18. Add label to the vector.
19. End for
20. Return labeled classes

Algorithm 1, Rough Set Based Features Selection Algorithm

4.2 Enhanced K-NN Classifier

In this work, the Enhanced KNN Classifier compares the test data set and train data set. In the data set, 70% is used for training and 30% for testing. It compares the performance ratio for both training and testing data set.

The basis of the proposed enhanced KNN algorithm is to assign membership using Gaussian membership function as a function of the data values distance from its K-nearest neighbours and the memberships in the possible classes. This enhanced KNN algorithm assigns class membership to the given vector rather than assigning the vector to a particular class. The advantage is that no arbitrary assignments are made by the algorithm. In addition, the vector's membership values provide improvement classification accuracy. Algorithm 2, explains the steps of the proposed Enhanced KNN algorithm.

Input: Set of 41 features from KDD data set
Output: A reduced set of six attributes Si.
Step1 : Select the all attributes
D(A_1, A_2,…….,A_n)Data set ,
δ – selected threshold.
step2: **For** i = 1 to n **do**
Step3: Select and remove the redundant columns
Step4: End for
Step5: **for** i = 1 to n **do**
Step6: Calculate threshold δ
Step7: Fuzzy classes $\{\mu\overline{P}(F_i), \mu\,\underline{P}\,(F_i)\}$
Step8: Calculate dependency γ_c
Step9: **End for**
Step10: **for** F_i of max γ_c **do**

Step11: **if**($\gamma_c > \delta$) $S_i = F_i$
Step12: **else if**($\gamma_c = 0$) $I_i = F_i$
Step13: **else if**($\gamma_c > \delta$) $D_i = F_i$
Step14: **end for**
Step15: Return S
Algorithm 2, Enhanced KNN Algorithm

5. Analysis of Experimental Results

5.1 Calculating Reduced Feature Set

Using the positive regions, the dependencies (γ) of the attributes with respect to the decision attribute is calculated in this work. Finally, the attributes with the maximum dependency value are taken and considered for further calculation and this iteration continues until the minimum set of attributes are found out. Table 2, shows the dependency values of the attributes from data set.

Table 2, Dependency Values of the Attribute

Chosen Attributes	Dependency Value
Choosing {4}	0.6285
Choosing {4,22}	0.80425
Choosing {4,22,31}	0.965
Choosing {4,22,31,37}	0.965
Choosing [4,22,31,,36,37]	0.9975
Choosing {4,5,22,31,36,37}	1.0
The Selected Attributes are {4,5,22,31,36,37,41}	

The calculation denotes the values of the dependency that is based on the concept that the attribute with the maximum dependency is selected. Our Proposed work reduces the computational time for detecting the attacks. The proposed rough set based feature selection algorithm reduces the number of attributes by selecting only the significant attributes. Significance is ranked on the basis of dependency value. Maximum of dependency higher the significance value and those attributes are added to the significant features. Table 3, shows the names of the selected attributes.

Table 3, Name of the Selected Attributes

S.No	Name of Selected Attributes
1	Dst_bytes
2	Flag
3	Count
4	Dst_host_count
5	Dst-host_srv_diff
6	Dst_host_srv_serror_rate

5.2 Performance of Enhanced Classifier

In this work, we have used the Enhanced- KNN for effective classification of the data set. Moreover, KNN is a non-parametric lazy learning algorithm and hence it does not make any assumptions on the underlying data distribution. Table 4, shows the detection rate provided by three classification algorithms namely SVM, KNN and M-KNN.

Table 4, Detection Rate with Full Set of Features

Exp. No.	SVM			KNN			M-KNN		
	Probe	DoS	Others	Probe	DoS	Others	Probe	DoS	Others
1	90.15	91.47	54.52	94.78	93.56	58.13	98.10	98.58	69.59
2	89.17	90.14	56.45	95.45	93.47	57.21	98.90	98.14	68.31
3	89.25	91.52	55.95	95.25	94.23	58.05	98.72	98.25	69.15
4	90.15	90.63	55.41	96.14	94.14	57.89	98.25	97.45	69.17
5	90.28	91.15	56.33	96.78	94.27	58.10	98.15	97.07	68.78

From table 4, it is observed that the detection rate is high when M-KNN is applied for classification. This is due to the fact that Enhanced KNN uses fuzzy rules derived from Gaussian membership function which is used for decision making along with distance measure. When it is compared with SVM, the performance of M-KNN is more significant. Similarly, the performance of KNN is not sufficient to reduce the false alarm rate. Table 5, shows the detection rate obtained from five experiments carried out with 10,000 records of the data set for each experiment on three classifiers namely SVM, KNN and M-KNN having same percentage of training data set for all the three algorithms. Similarly, 30% of the data set were taken for performing the testing.

Table 5. Detection Rate with Reduced Set of Features

Exp. No.	SVM			KNN			M-KNN		
	Probe	DoS	Others	Probe	DoS	Others	Probe	DoS	Others
1	92.13	93.30	60.73	96.20	96.00	63.47	99.51	99.29	71.62
2	91.78	92.15	61.10	97.55	96.24	65.05	99.25	99.45	69.42
3	92.67	93.20	60.92	97.25	96.25	64.60	99.14	99.75	74.32
4	92.29	93.18	61.20	98.30	95.98	64.72	99.13	99.23	73.43
5	92.23	93.90	62.43	98.50	96.89	63.20	99.18	99.24	71.17

From table 5, it is observed that the detection rate is increased when selected features are used. This is due to the fact that the fuzzy rules applied on reduced features set have equivalent crisp set rules. Moreover, there no contradicting attributes which made the decision process easy for all the three algorithms namely SVM, KNN and M-KNN.

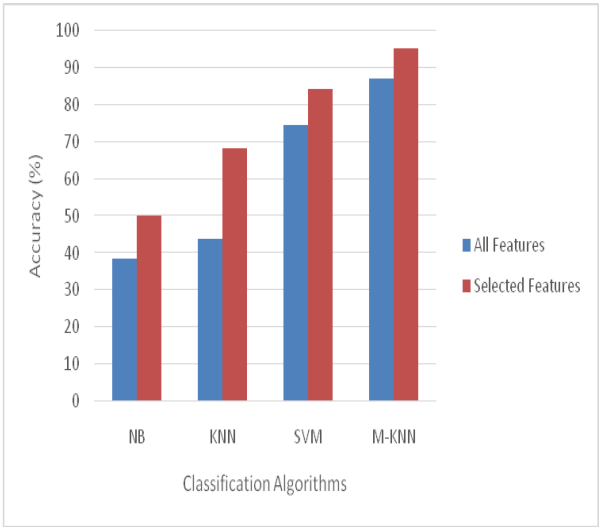

Figure. 2. Accuracy Analysis for Classifiers

Figure 2, shows the accuracy analysis for the four classifiers namely Naïve Bayes classifier, SVM, KNN and Enhanced KNN. From this figure, two observations can be made. First, the selected set of features increases the classification accuracy. This is due to fact that selected features take only the necessary rules for decision making. Therefore, the classifier is not confused, leading to increase in accuracy. Second, Enhanced KNN out performs all the other algorithms. This is because Enhanced rules are overcoming the boundaries in decision making.

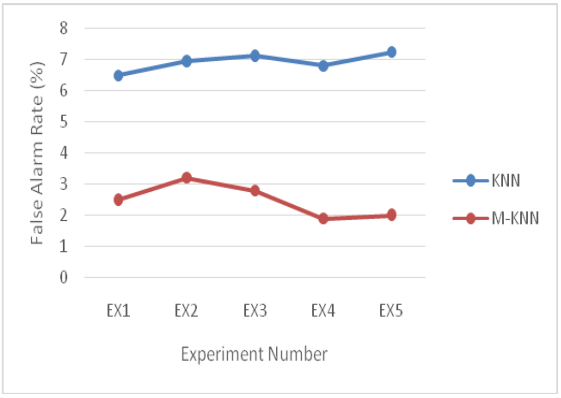

Figure 3, False Alarm Rate Analysis

Figure 3, shows the false alarm rate analysis for the two classifiers namely KNN and Enhanced KNN. From figure 3, it is observed that enhanced KNN has less false alarm rate in comparison with KNN. This is because; in all the five experiments fuzzy rules were applied for conflict resolution.

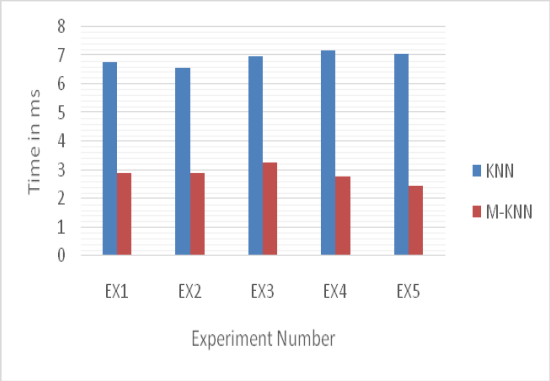

Figure 4, Time Analysis

Figure 4, shows the time analysis for the two classifiers namely KNN and enhanced KNN. From figure 4, it is observed that enhanced KNN takes less time in comparison with KNN. This is because in all the five experiments fuzzy rules were applied for fast convergence.

6. Conclusion

In this proposed work, a new IDS has been developed and implemented with a simple feature selection algorithm. The work was carried out on selected attributes computing the result with 41 attributes seems to be very complicated and difficult to achieve accuracy. The proposed work selects only the significant features that have the high probability of predictive measure. With the reduced set, we have reduced the computational time. Further, the Enhanced K-NN classifier helped in achieving the greater accuracy. Hence, we computed the result in an efficient manner to prevent the attacks that improves the security.

References

[1] Kristin Bennett, Jinbo Bi, Mark Embrechts, Curt Breneman and Minghu Song, "Dimensionality Reduction via Sparse Support Vector Machines " , *Journal of Machine Learning Research,* vol.3, pp1229-1243, 2003.

[2] M. Ramaswami, R. Bhaskaran, "A Study on Feature Selection Techniques in Educational Data Mining", *Journal of Computing,* Vol.1, No.1, pp 7-11, 2009.

[3] XiuboGeng, Tie-Yan Liu, Tao Qin and Hang Li, "Feature Selection for Ranking", *SIGIR '07 Proceedings Of The 30th Annual International ACM SIGIR Conference On Research And Development In Information Retrieval,* pp 407-414 , 2007.

[4] Jennifer G. Dy and Carla E. Brodley, "Feature Selection for Unsupervised Learning", *The Journal of Machine Learning Research archive* , vol. 5, pp 845-889 ,2004.

[5] George Forman, "An Extensive Empirical Study ofFeature Selection Metrics for Text Classification", *Journal of Machine Learning Research,* vol.3,pp1289-1305, 2003.

[6] Lei Yu Huan Liu, "Efficient Feature Selection via Analysis of Relevance and Redundancy", *Journal of Machine Learning Research,* vol.5 pp1205–1224, 2004.

[7] R.GeethaRamani, S.SivaSathya, Sivaselvi K. "Discriminant Analysis based Feature Selection in KDD Intrusion Dataset", *International Journal of Computer Application,* vol.31, no.11, 2011.

[8] Wei-ZhiWua, and Wen-XiuZhanga, "Constructive and Axiomatic Approaches of Fuzzy Approximation Operators", *Information Sciences,* vol.159, no. 3, pp 233–254, 2004.

[9] Richard Jensena, Andrew Tusonb and Qiang Shena, "Finding Rough and Fuzzy-Rough Set Reduces with SAT", Information Sciences, vol. 255, pp 100–120, 2014.

[10]Sindhu, S,Geetha, S. and Kannan, A. "Decision Boundary based Light Weight Intrusion Detection using a Wrapper Approach", *Expert Systems with Applications,* vol.39, pp.129-141, 2012.

[11]Iosif-ViorelOnut and Ali A. Ghorbani, "A Feature Classification Scheme for Network Intrusion Detection", *International Journal of Network Security,* Vol.5, no.1, pp.1–15, July 2007.

[12]Sannasi Ganapathy, Kanagasabai Kulothungan, Sannasy Muthurajkumar, Muthusamy Vijayalakshmi, Palanichamy Yogesh and Arputharaj Kannan, "Intelligent Feature Selection and Classification Techniques for Intrusion Detection in Networks: a Survey", *EURASIP Journal on Wireless Communications and Networking,* vol.271 , pp 1-16,2013.

[13] Huaguang, Zhang, Jinhai, Liu, Dazhong, Ma and Zhanshan Wang, 'Data-Core-Based Fuzzy Min–Max Neural Network for Pattern Classification', *IEEE Transactions on Neural Networks,* vol. 22, no. 12, pp. 2339-2352, 2011.

[14]Debar, H., Becker, M. and Siboni, D. "A Neural Network Component for an Intrusion Detection System", *IEEE Symposium on Research in Computer Security and Privacy,* pp.240-250, 2012.

[15]Ahmed Patel, Mona Taghavi, Kaveh Bakhtiya, Joaquim and Celetino Junior, "An intrusion detection and prevention system in cloud computing: A systematic review", *Journal of Network and Computer Applications, vol.36,* no.1, pp 25-41, 2013.

[16]Moradi M and Zulkernine M "A Neural Network based System for Intrusion Detection and classification of Attacks", *Proceedings of IEEE International Conference on Advances in Intelligent Systems-Theory and Applications,* vol.148, pp.1-6, 2011.

[17]SarasammaS, Zhu, Q. and Huff, J."HierarchicalKohonen Net for Anomaly Detection in Network Security", *IEEE Transactions on System, Man, Cybernetics, Part Cybernetics, vol.35, No.2, pp.302*-312, 2005.

[18]Wang Jianping, Chen Min and Wu Xianwen, "A Novel Network Attack Audit System based on Multi-Agent Technology", *Physics Procedia, Elsevier,* Vol 25, pp.2152-2157, 2012.

[19]SnehalA.Mulay, Devale, P.P. and Garje, G.V. "Intrusion Detection System using Support Vector Machine and Decision Tree", *International Journal of Computer Applications,* Vol.3, pp-975-987, 2010.

[20]SnehitaModi, Paramjeet Singh, Shaveta Rani, "Performance Improvement of Mobile Ad hoc Networks under Jamming Attack", *International Journal of Computer Science and Information Technologies,* Vol. 5 No. 4, pp.5197-5200, 2014.

[21]Mohammad SazzadulHoque, Md. Abdul Mukit, and Md. Abu NaserBikas, "An Implementation of Intrusion Detection System Using Genetic Algorithm", *International Journal of Network Security & Its Applications,* Vol.4, No.2, 2012.

[22]Kartit A,Saidi A, Bezzazi F, Marraki ME andRadi A, "A New Approach to Intrusion Detection System", *Journal of Theoretical and Applied Information Technology,* vol. 36, pp284–9 , 2012.

[23]Chun-WeiTsai,"Incrementalparticle swarm optimisation for intrusion detection",Networks,*IET*,Vol.2, Issue:3,pp.124–130,26August 2013.

[24]JohnM.Fossaceca,ThomasA.Mazzuchi and ShahramSarkani,"MARK- ELM:Application OfANovel MultipleKernelLearningFrameworkFor ImprovingTheRobustnessOfNetworkIntrusion Detection",*Expert SystemswithApplications,*Vol.42,Issue8,pp.4062–4080,15May2015.

[25]RiyanatShittu,AlexHealingRobert,Ghanea-Hercock, RobinBloomfield andMuttukrishnanRajarajan,"IntrusionAlertPrioritisationAndAttack DetectionUsingPost-CorrelationAnalysis" ,*Computers&Security,*Vol.50, pp.1-15,May2015.

[26]Adel Sabry Eesa,Zeynep Orman and Adnan Mohsin Abdulazeez Brifcani, "A Novel Feature-Selection Approach Based On The Cuttlefish Optimization Algorithm For Intrusion Detection Systems", *Expert Systems withApplications,*Vol.42, Issue5,pp.2670–2679,1April2015.

[27]Abduvaliyev, A.Pathan, A.-S.K.JianyingZhou, "Onthe V i t a l A r e a s o f Intrusion Detection Systems in Wireless Sensor Networks", *CommunicationsSurveys&Tutorials,* Vol.15, pp 1223–1237,31July2013.

[28]YunhoLee1,Sang-GuunYoo2andSoojinLee,"AnEfficientDetection AndManagementOfFalseAccusationAttacksInHierarchicalAd-Hoc Networks", *KSIITransactionsonInternetand InformationSystems*,vol 6,no.7, July2012.

[29]Thongkanchorn,K. Ngamsuriyaroj,S. Visoottiviseth,V. "Evaluation StudiesofThree Intrusion Detection Systems under Various AttacksandRule Sets",*TENCON2013-2013IEEE Region 10Conference,* pp.1–4,22-25Oct.2013.

[30] Nader P, Honeine P and Beauseroy," -Normsin One-Class Classificationfor Intrusion Detectionin SCADA Systems", *Industrial Informatics, IEEE Transactions,* vol,10,Issue:4,pp.2308–2317,30June2014.