Indonesian Journal of Electrical Engineering and Computer Science
Vol. 14, No. 1, April 2019, pp. 101~112
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v14.i1.pp101-112

A novel approach for selective feature mechanism for two-phase intrusion detection system

B Narendra Kumar[1], M S V Sivarama Bhadri Raju[2], B Vishnu Vardhan[3]
[1]Department of Computer Science Engineering, Sri Sai Jyothi Engineering College, Hyderabad, Telangana, India
[2]Department of Computer Science Engineering, SRKR Engineering College, Bhimavaram, Andhra Pradesh, India
[3]Department of Computer Science Engineering, JNTU College of Engineering Manthani, Peddapalli, Telangana, India

Article Info

Article history:

Received May 8, 2018
Revised Nov 7, 2018
Accepted Jan 17, 2019

Keywords:

Accuracy
Correlation
IDS
MI
RNN
SVM

ABSTRACT

Intrusion Detection is an important aspect to secure the computing systems from different intrusions. To improve the accuracy and to reduce the computational time, this paper proposes a two-phase hybrid method based on the SVM and RNN. In addition, this paper also had a proposal to obtain a few sets of features with a feature selection technique in which the detection performance increases. For the two-phase system, two different feature selection techniques were proposed which solves both the linear dependency and non-linear dependency between the features. In the first phase, the RNN combines with the proposed Joint Mutual Information Maximization (JMIM) based feature selection and in the second phase, the Support Vector Machine (SVM) combines with correlation based feature selection. Extensive simulations are carried out over the proposed system using two different datasets, NSL-KDD and Kyoto2006+. The performance is measured through the evolution metrics such as Detection Rate (DR), Precision, False Alarm Rate (FAR), Accuracy and F-Score. Furthermore, a comparative analysis with few recent hybrid frameworks is also enumerated. The obtained results signify the effectiveness of proposed method.

Corresponding Author:

B.Narendra Kumar,
Department of Computer Science Engineering,
Jawaharlal Nehru Technological University,
Hyderabad, Telangana, India.
Email: bnkphd@gmail.com

1. INTRODUCTION

The development of computer networks, particularly the internet has brought substantial changes in the daily lives, flexibilities in the business relations, organizations in the services provision etc. Along with these conveniences, it also brought various security threats which have become a serious concern due to the constant appearance of new susceptibilities and attacks. Hence there is a need to develop a more efficient secure strategies which protects the systems form these threats and also maintains the data confidentiality, integrity and availability. IDS is one of such securing strategies which have gained a lot of popularity due to its flexibility in the detection and prevention of different known and unknown security threats [1], [2]. An IDS monitors the events and collects network packets in a computing architecture. By analyzing the packets acquired form the system, the IDS detects abnormal patterns and block those malicious connections from intruders or attackers. In the last decade, the research over the IDS has obtained a lot of attention from various researchers [3], [4].

Depending on the methodology, the intrusion detection approaches are categorized as anomaly based detections and misuse based detections. An anomaly based detection techniques identify the attacks based on their behavior. For a connection, whenever a deviation is observed from the normal behaviors, it is classified as attack [5]. Just because of this concept, the anomaly based detection is considered as a

classification problem. On the other hand, the misuse based detection detects an intrusion by matching it with predefined signatures. Hence to build the misuse based detection system, there is a necessity of knowing the profiles of attacks. The main drawback of the misuse based detection is high FAR in the case of an unknown attacks. For unknown attacks, the misuse based detection never identifies because, the profiles of such attacks are not known to the system. However the anomaly based detection is identify such types of anomalies.

Deviating from the main objective, i.e., the maximum detection accuracy, computational time is also an important factor which plays asignificant role in the performance analysis of IDS [6], [7]. The computational time is measured as the time taken by the IDS to detect the attack and reflected with the features of dataset. There were so many approaches that are developed by focusing towards the reduction of feature count from the dataset which has an impact on the computation time. Based on the methodology developed for feature analysis, the earlier developed approaches were categorized as classifier dependent and classifier independent. Though there are so many approaches, still there is a scope to improve the performance of IDS.

This paper proposes a novel hybrid intrusion detection framework to achieve an improved detection performance in the detection of different attacks in the IDS. The novelty of the proposed approach lies at the preprocessing phase at which the optimal set of features are selected by a perfect discrimination between different attacks. Considering the class relevancy with feature, this method proposes a new MI Based Feature Selection Mechanism. Furthermore this approach also combines two machine learning algorithms such as Multi-Class SVM and RNN for the anomaly detection and misuse detection respectively. The most popular NSL-KDD and Kyoto2006+ for intrusion detection datasets were used to simulate the proposed approach and the performance is measured through computational time.

Remaining paper is organized as follows; Section II is for the illustration of Literature survey. Section III describes the preliminaries of proposed approach. Section IV illustrates the complete details of proposed methodology. The experimental results are conducted in Section V and finally the conclusions are given in section VI.

2. LITERATURE SURVEY

2.1. Feature Selection (FS) approaches

FS is used as a main aspect in different applications relevant to intelligent and expert systems such as machine learning, data mining, anomaly detection, image processing, natural language processing and bio-informatics. FS is generally accomplished over the data before training it to the classifier. This process of FS is also termed as variable subset selection, feature reduction or variable selection. For IDS, features are more important which makes the system robust for any circumstances. Basically the feature selection methods are classified into two classes; classifier dependent and classifier independent. Further the classifier dependent approaches are classified as wrapper and embedded methods.Compared to the classifier dependent methods, the classifier independent methods are computationally efficient and more scalable, in terms of data dimensionality and from classifier independence. Pearson Correlation Coefficient (PCC) [12], Fisher's Discriminate Ratio (F-Score) [8], MI [9], Rough set theory [10], and Data Envelopment Analysis [11] are some of the filter based feature selection approaches. Among these approaches MI gained an increased popularity due to its independent nature towards the data type includes numerical and categorical with two or more class values. Further the MI doesnot makes the assumption of linearity between the variables.

Beauqier and Hu [12] developed IDS by combining different methods like "Pearson's Correlation coefficients-Rank (PCC-R)", in which the PCC-R was accomplished for the evaluation of Euclidean distances between various methods such as "Probabilistic Finite State Automata (PFSA)" and "Naive Bayes, Bayes one-step Markov model". Though the combination of these methods achieves effective results, the FAR is observed to be high. Jin et al. [13] utilized covariance matrix of sequential samples to detect multiple network attacks. Akshadeep et.al., [14] considered the information gain and correlation for FS and used artificial neural network for classifying the attacks in the IDS. This method mainly focused towards the less occurring and frequent occurring attacks. Chaouki Khammassi and Saoussen [15] developed a three stage IDS, the three stages are preprocessing, FS and classification. GA-LR wrapper is accomplished for FS and the three decision tress classifiers are used for classification.A new method is proposed in [16] to solve the many-objective problem to select the optimal feature set in the IDS. This strategy is based on two methodologies such as "predefined multiple targeted search" and a "special domination method". Here the first method is considered for population evolution. Based on the proposed aspects, the NSGA-III is accomplished to extract an adequate set of features to achieve an effective performance. Further an "improved NSGA-III (I-NSGA-III)" is also developed based on the process of niche preservation [24].

Amiri [17] developed two distinct FS approaches to extract the optimal feature set and they are compared with MI based FS method. A new metric which evaluates the feature goodness is accomplished in

this approach. Both the linear and non-linear measures are accomplished in this method to extract the optimal feature in all directions. Further this approach used the LS-SVM to construct the IDS.By extending the MI, a new filter basedFS mechanism is proposed by Mohammad M. Ambusaidi [22] to perform IDS. A MI based FSalgorithm is developedthat methodicallychooses the optimal features for classification. The proposed FS by [22] solves both linear and non-linear dependencies between the variables and tries to select an optimal feature set by which the primary objective of IDS is achieved. Further an IDS, named as "Least Square Support Vector Machine based IDS (LSSVM-IDS)", is built through the obtained features.

To reduce the FAR followed by the computational time in the IDS, Sumiya et.al., [18-20] studied different FS methodologies along with SVM to build a hybrid IDS models. Recently, in [19], the IDS model is accomplished through Chi-Square based FS and the MC-SVM. Here the Chi-square is used for FS and MC-SVM is for classification of different classes. Though the Chi-square method is a simple FS technique, it didn't illustrate the dependencies between the variables by which the entire feature set need to be scanned for every connection. This process increases the computational time. An IDS was built by Saxena and Richariya [21] using the Information gain ratio, SVM and PSO. Though the accomplishment of PSO achieves a higher accuracy levels, it did not focused on the evaluation of computational complexity, which is an effective factor in the IDS performance.

2.2. Hybrid Approaches

Though there are so many approaches developed using different machine learning algorithms, the performance of an IDSis further increased by adopting two different classifiers, one for anomaly and another for misuse. In the case of anomaly, classification process is easy compared to the misuse because, the anomaly based detection focuses to classify normal and abnormal classes only whereas the misuse based detection has a typical process to classify more classes. Hence a new class of IDS approaches called as hybrid approaches are developed by combining two classifiers to perform anomaly and misuse detection tasks individually. Different methods are developed by different authors by combining different classifiers like SVM and decision tree [31], "k-means and k-NN" [32], SVM and ANN [33] etc.

Abdulla Amin Aburomman and Mamun Bin [23] focused to combine two classifiers, SVM and K-NN. Totallyanensemble of six SVM and six k-NN classifiers are used.PSO and meta-PSO are the two meta-heuristic algorithms which were used to create these ensembles. To acquire a detailed knowledge about the detection of network Intrusions, S. Y. Ji et.al., [25] designed a network intrusion detection through a multi-level strategy. Mainly this strategy composed of three phases, (1) to study the detailed analysis and to know the abnormalities in in the network traffic, a set of reliable rules are created, (2) generation of a extrapolative model to observe the perfect attack strategies, and (3) Integration of a graphicinvestigation tool to perform an interactive graphicinvestigation and to validate the intrusions recognizedwith obvious reasons [25] Accomplished decision tree [28], SVM, neural network algorithms as classifiers in the multi-level fashion. In [26], an effective IDS framework was designed based on the "Time-Varying Chaos Particle Swarm Optimization (TVCPSO)". TVCPSO is accomplished here for the concurrentFS and for the parameter setting. The FS is carried out here through the "Multiple Criteria Linear Programming (MCLP)" and classification is through SVM. A New Objective Function is provided in the developed methods to provide a trade-off between the minimization of FAR and maximization of DR, along with the number of features.

Further to achieve an optimal performance in the IDS, Wathiq et.al., [27] combined the two Machine Learning (ML) algorithms, SVM and CNN. Animproved k-means algorithm is also accomplished to reduce the size of dataset and a multi-layered prototype is proposed to increase the DR. To improvise the performance of classifier for the IDSs, a novel supervised learning algorithm assisted to the semi supervised learning algorithm with fuzziness is proposed by Rana Amir et.al., [29] utilizing the unlabeled test samples. Here to get the fuzzy membership vector as output, a "Single hidden Layer Feed-forward Neural network (SLFN)" is trained. The categorization of samples like High fuzziness, medium fuzziness and low fuzziness over the unlabeled samples is done through the fuzzy quantity. Again the classifier is trained after including the respective category into the respective connection in training set. "Optimum Path Forest (OPF)" is a graph based ML algorithm which was developed to overcome some problems with the conventional ML algorithms. Based on the OPF, H Bostani and M Sheikhan [30] proposed an IDS through animproved OPF to increase the performance of conventional OPF w.r.t the FAR, DR and the execution time. Further to achieve the scalability in the large size datasets, [30] also employed the "k-means clustering", as a segregatingunit. Recently, to achieve both benefits with respect to the DR and computational time, a selective feature based hybrid framework is proposed by B.Narendra et.al., [42] by combining the SVM classifier and the Convolutional Neural Network. An extended MIFS is proposed to detect the anomaly and the PLCC is used to detect the misuse.

3. PRELIMINARIES

3.1. Feature Selection (FS)

FS is asignificant aspect in the IDS. There were so many approaches that are developed to achieve an efficient performance in the IDS. Among the earlier developed FS methodologies, MI based FS approaches hasproved a more significant results in the detection performance. The MI was first developed by Battiti in 1994 [9], also known as a first order incremental search algorithm. Battiti proposed MI to select more RelevantFeatures (RF) from the initial set of 'N' features. Instead of evaluating the JMI between theClass Label (CL) and the Selected Features (SF), Battiti's MI evaluates MI between the CL and Candidate Feature (CF), relationship between the CF and the already SFs. Further there are many variants that are proposed based on MI such as MIFS-U [36], mRMR [37], NMIFS [38], MIFS-ND [39] and JMI [40]. Among these methods, the MIFS-ND calculates the MI between theCL and theCF in the context of SF subset. MIFS-ND accomplished a Genetic Algorithm (GA) to select an optimal feature which maximizes the MI with CL and minimizes the MI with the remaining SFs. Further some more methods are also developed based on the MI. However the following drawbacks are observed with the earlier developed approaches.

a) Class Irrelevancy

In the afore said methods, the redundancy is measured based on the MI value between the CF and features in the SF subset, but never considered the CL. If the MI between the CF and SF in the subset is less, then the CFs is considered as redundant features, but this phenomenon is wrong when the redundant candidate features share different information with another class.

b) Over Estimation of feature significance

In the case of high correlation of candidate feature with one or some pre-selected features, the candidate feature is assumed to share more information about the features selected in the subset, but at the same time the candidate feature can be an independent feature form the majority features in the selected feature subset. In that condition, the value of objective function is greatin spite of the redundancy of the CFand to some features within the subset. This problem was occurred in the methods like MIFS-U, mRMR, MNIFS, MIFS-ND which follows a forward search mechanism and a cumulative summation to estimate the solution.

3.1.1. Joint Mutual Information Maximization (JMIM)

In this study, a new FS method is proposed based on the MIFS, named asJMI Maximization (JMIM). JMIM is a combined form the JMI and Maximum of the Minimum (MIM). JMIM is aimed to address the above problems, class irrelevancy and the Overestimation if feature significance, which ensures when the cumulative summation is accomplished.

The FS process is in such a way that for a given full feature set F of size N, it needs to select a feature subset, S, S⊆F, with dimensions K, K≤N, by which the classification accuracy is equal or highwhen compared it to the accuracy obtained through the full set of features, F. Simply it can also be defined as a FS that extracts the features which have maximum MI with the CL, i.e., I(S; C).

Based on these aspects, the feature relevance is defined as, for an already selected feature subset, S, he feature f_i is said to be more relevant than the feature f_j if the MI between f_i and S with respect to the class C $(I(f_i, S; C))$ is greater than the MI between the feature f_j and S with respect to the class C $\left(I(f_j, S; C)\right)$, simply, $(I(f_i, S; C)) > \left(I(f_j, S; C)\right)$.

Further, the feature relevance defines through the Joint MI. Let F be the full set of features, S be the subset of features which was already selected for the Feature set F. Let a feature f_i, $f_i \in F - S$, and $f_s \in S$, the m-Joint MI is defined as the MI between f_i and the features present in the already selected feature subset S. The minimum value of m-Joint MI is referred as minimum joint MI, i.e., $\min_{s=1,2,\ldots,K} I(f_i, f_s; C)$. A larger value of joint MI of f_i and the features in the subset S denotes a high relevance with the class label C. Further a larger value of joint MI also denotes that the m-joint MI of other features, f_j, f_i and $f_j \in F - S$ denotes the minimum joint MI between the features f_j and f_i . Simply it denotes that, compared to the feature f_i, the feature f_j, shares less information towards the class label C. According to the above definitions, the feature which shares maximum information is said to be more relevant.

Further a new definition is given for redundancy from the given set of features F, and a selected feature subset S, a feature f_i is said to redundant to the selected feature subset S if it doesnot share new information with the class C. If the feature f_i is highly correlated with a feature f_s, $f_s \in S$, then the probability of mass functions of f_i, f_s and (f_i, f_s) are equal, i.e., $P(f_i) \cong P(f_s) \cong P(f_i, f_s)$.

Based on the above discussions, to overcome the problem of over estimation feature significance, this work accomplished a new method called JMIM to select the optimal feature set by which the accuracy

increase with less number of features extracted from dataset. It is a combined form of Joint MI and MIM, through which the most RFs are chosen. The new criterion for the FS according to the JMIM is formulated as

$$f_{JMIM} = \arg \max_{f_i \in F-S} \left(\min_{f_s \in S} (I(f_i, f_s; C)) \right) \quad (1)$$

Where:

$$I(f_i, f_s; C) = I(f_s; C) + I\left(f_i; {}^{C}/_{f_s}\right) \quad (2)$$

$$I(f_i, f_s; C) = H(C) - H\left({}^{C}/_{f_i}, f_s\right) \quad (3)$$

$$I(f_i, f_s; C) = \left[-\sum_{c \in C} p(c) \log(p(c))\right] - \left[\sum_{c \in C} \sum_{f_i \in F-S} \sum_{f_s \in S} log\left(\frac{p(f_i, f_s, {}^{C}/_{f_s})}{p({}^{f_i}/_{f_s}) p({}^{C}/_{f_s})}\right)\right] \quad (4)$$

This method follows the forward search mechanism in the iterative fashion to find the subset of most RFs of size *k* from the original full feature set.

3.2. Classification

Once the features are extracted, they are processed for classification and here two algorithms are accomplished for classification, they are namely Recurrent Neural Network (RNN) and Multi-class SVM. RNN is an extended version of the most popular Feed forward Neural network [34]. Instead of linear connections in the feed forward neural network, the RNN has cyclic connections which make it most powerful to solve the problems in the linear and non-linear sequences. To train the RNN, generally the Back Propagation Through Time (BPTT) is accomplished. However, the common drawback of the basic RNN is exploding gradient s and vanishing gradients. To overcome these issues, a "Long Short-Term Memory (LSTM)" based RNN is introduced previously [35]. Here the LSTM-RNN is for the classification of normal class from attack classes and Multi-class SVM is for further individual classification.

4. PROPOSED SYSTEM

This paper proposes a new hybrid IDS framework by combining the LSTM-RNN and Multi-Class SVM [41]. The complete system is developed under two phases, anomaly detection and misuse detection. Under anomaly detection, this work accomplishes the JMIM based FS mechanism for FS and the obtained MI data, the LSTM–RNN is accomplished to classify the data into attack and normal classes. In the second phase, this work tries to classify the abnormal/attack classes into various types such as DoS, Probe, U2R, and R2L by using Multi-class SVM. In the second phase, the Pearson correlation coefficient is used for FS technique. An overall schematic of developed IDS framework is depicted in Figure 1.

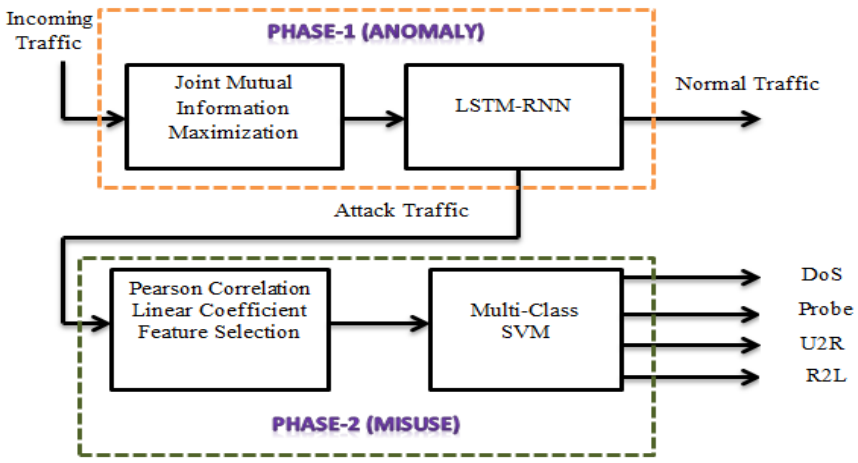

Figure 1. Overall architecture of proposed framework

According to the standard IDS, initially the incoming traffic is preprocessed to make it compatible with the system characteristics. Under the preprocessing, whenever the data is in more than one format, then it is processed for data normalization. For example, the "NSL-KDD" data is having both categorical and numerical formats. In the data normalization, the categorical data is also converted into the numerical format because the system accepts only the numerical data. Further the preprocessing phase accomplishes the feature extraction or FS mechanism over the normalized data to obtain only a few set of feature form initial full set. This is mainly to remove the redundant data by which the unnecessary computation burden arises. In this paper, two filter based FS techniques are accomplished to obtain a few and efficient set of features by which the main objective of IDSsuch as Accuracy, DR, and precision will be achieved more efficiently. Firstly, the JMIM based FS is accomplished to find the probability of occurrences of normal over abnormal connections followed by abnormal over normal connections. Based on the obtained JMIM data, few efficient features are extracted for both normal and abnormal connections by which the overall normal/abnormal connections can be represented without any data loss. The obtained feature set which describes the complete normal and attack connections are depictedin the Table 1.

Table 1. Obtained optimal feature set results through JMIM

Class	Feature Count	Feature set
Normal	11	$f_2, f_3, f_4, f_5, f_{22}, f_{24}, f_{25}, f_{31}, f_{32}, f_{36}, f_{37}$
Attack	17	$f_1, f_2, f_3, f_4, f_5, f_{12}, f_{15}, f_{17}, f_{23}, f_{24}, f_{27}, f_{29}, f_{32}, f_{33}, f_{35}, f_{37}, f_{40}$

After obtaining the few set of features for normal and attacks connections, they are trained through the LSTM-RNN algorithm. Here LSTM-RNN algorithm is a supervised deep learning algorithm which considers the previous state to predict the resent state which makes the system to classify more accurately. For example, whenever the i^{th} connection is classified as attack connection, that status was stored in the memory of LSTM to consider it as feedback for the next connection prediction.

In the next phase, the connections which are classified as attacks are processed for misuse detection. In this phase, the PLCC is used to extract the optimal set of feature based on their correlations. There exist linear and non-linear relations when the entire dataset is consideredhence the proposed system accomplished a MI based technique which extracts both relations and by which the classification becomes more flexible. In the case on misuse detection, the entire attack connections are assuming to be linearly dependent and the proposed system extracts those linear relations with PLCC, a technique in the extraction of linear dependency between the variables. Based on the evaluated correlations between the attack connections, the features which are maximally correlated with all attacks are only considered as efficient feature set and only those features are trained to the system. The obtained few sets of features of attacks are depicted in the Table 2.

Table 2. Obtained optimal feature set through PLCC

Class	Feature Count	Feature set
DoS	10	$f_2, f_3, f_4, f_6, f_8, f_{12}, f_{23}, f_{24}, f_{32}, f_{36}$
Probe	12	$f_2, f_3, f_4, f_5, f_{17}, f_{24}, f_{27}, f_{29}, f_{32}, f_{35}, f_{36}, f_{40}$
U2R	13	$f_1, f_2, f_3, f_4, f_6, f_{12}, f_{16}, f_{23}, f_{24}, f_{32}, f_{33}, f_{34}, f_{36}$
R2L	10	$f_1, f_3, f_5, f_9, f_{10}, f_{11}, f_{22}, f_{24}, f_{32}, f_{36},$

After extracting the optimal feature set thorough PLCC, they are trained through Multi-Class SVM classifier. Since the SVM is a binary classifier, the accomplishment of SVM is carried out at multiple levels, hence named as multi-class SVM. Initially the entire attack traffic is classified as DoS and the remaining (Probe, U2R and R2L) through the SVM classifier 1. Further the SVM classifier 2 classifies the remaining attacks into two classes such as probe and remaining (U2R and R2L). Finally the SVM classifier 3 classifies the remaining traffic into the U2R and R2L classes. Totally, the number of SVM classifiers required to accomplish the misuse detection are three.

5. EXPERIMENTAL RESULTS AND ANALYSIS

5.1. Dataset

In the IDS field, there are only few publicly available datasets to evaluate the performance of IDSs. "KDD cup 99 data set" is a most famous and comprehensive intrusion detection data set. It consists of totally 5 classes among which normal is one class and the remaining classes are attacks (DoS, Probe, U2R

and R2L). It contains approximately five million training records and two million testing records. Every record of this dataset is formulated with 41 different features (both qualitative and quantitative). Each and every record is labeled as either attack or normal. Among these 41 features, 36 features are continuous, three are symbolic features and two are binary features. Since the most of the classifiers accepts the numerical values, the symbolic values can be converted into numeric values.

Further the "NSL-KDD dataset" [43] is a more proliferative dataset which was derived from the most familiar "KDD cup 99 dataset". It was extracted from the "KDD cup 99 dataset" after solving some intrinsiccomplications existing in itlike redundant records in a huge number. It consists of one training set, KDDTrain+ and two testing sets, KDDTest+ and KDDTest^{-21}."The complete details of "NLS-KDD dataset" are illustrated in Table 3.

Table 3. Details of NSL-KDD dataset

Dataset	Normal	DoS	Probe	U2R	R2L	Total
KDDTrain+	67343	45927	11656	52	995	125973
KDDTrain_20%	13449	9234	2289	11	209	25192
KDDTest+	9711	7458	2421	200	2754	22544
KDDTest^{-21}	2152	4342	2402	200	2754	11850

Kyoto2006+ is one more dataset, introduced by Song et al.[44].This dataset consist of the following 24 statistical features; 14 conventionalfeatures and 10 additional features. Among them, the first 14 featureswere extracted based on KDD Cup 99 data set.Among 41 original features of KDD Cup 99 data set, only 14significant and essential features are extracted from the raw traffic data obtained by honeypotsystems that are deployed in Kyoto University. Addition to those 14 features, additionally 10 more features are also extracted which may enable the users to investigate more effectively what happens in the networks. For experimental analysis on Kyoto 2006+ dataset, the data of 12, 13, 14, 15 and 16 of November 2006 are selected. The total number of connections for the selected dataset is 93240. According to the 'Label' present in the Kyoto 2006+ dataset, the total number of connections recognized as attacks are 71885 and the total number of connections recognized as Normal are 21355. To test the proposed system, the total number of connections considered is 27972. Among these connections, the total number of normal connections are 6410 and the total number of attack connections are 21562. In the evaluation criteria, the performance metrics namely, Accuracy, Precision, FAR, DR, and F-Score [41] are considered to evaluate the performance of developed system.

5.2. Results

To assess the performance enhancement of the developed IDS framework, a sequence of tests were accompanied on the "NSL-KDD dataset and Kyoto2006+ dataset". All experiments were implemented in the MATLAB 2014b with hardware configuration of one Terabyte Hard Disk and eight Gigabyte RAM. Initially the training dataset is accomplished for preprocessing and then the obtained features are trained to the system Further the testing dataset was subjected to testing after completing the preprocessing over it. Since the NSL-KDD dataset consists of five different classes, the proposed hybrid framework classifies the total classes in two phases. In the case of Kyoto2006+ dataset, there are only two classes such as attack and normal. To test the Kyoto2006+ dataset, initially the training set was processed for anomaly detection and the obtained normal and attack connections are accomplished. Further it is again processed through misuse detection and the obtained results are accomplished. Based on these two observations, the overall performance is evaluated.

The obtained results after the accomplishment of proposed approach over the KDDTest+ dataset which are represented in Table 4 and Table 5. Table 5 represents the details of first phase and the second phase results are represented in Table 6. According to the proposed methodology, in the first phase, the connection is classified into normal and attacks only. For a given total 22544 connections, the first phase classified 12310 as attacks and 9671 as normal. Further the second phase classifies the 12310 attack connections into the respective class such as "Dos, Probe, U2R and R2L". The details are represented in Table 6, 7052 are DoS, 2285 are probe, 161 are U2R and 2522 are R2L.

Table 4. Confusion matrix obtained in the first phase over KDDTest+

		Predicted	
		Attack	Normal
Actual	Attack	12310	523
	Normal	40	9671

Table 5. Confusion matrix obtained in the Second phase over KDDTest+

	DoS	Probe	U2R	R2L	Total
Dos	**7052**	47	14	41	7154
Probe	21	**2285**	9	10	2325
U2R	15	07	**161**	07	190
R2L	59	38	22	**2522**	2641

The obtained results after the accomplishment of proposed approach over the KDDTest-21 dataset are represented in Table 6 and Table 7. Table 7 represents the details of first phase and the second phase results are represented in Table 8. According to the proposed methodology, in the first phase, the connection is classified into normal and attacks only. For a given total 11850 connections, the first phase classified 9599 as attacks and 2115 as normal. Further the second phase classifies the 9599 attack connections into the respective class such as Dos, Probe, U2R and R2L. The details are represented in Table 8, 4298 are DoS, 2378 are probe, 197 are U2R and 2726 are R2L.

Table 6. Confusion matrix obtained in the first phase over KDDTest^{-21}

		Predicted	
		Attack	Normal
Actual	Attack	9599	99
	Normal	37	2115

Table 7. Confusion matrix obtained in the Second phase over KDDTest^{-21}

	DoS	Probe	U2R	R2L	Total
Dos	**4193**	54	16	35	4298
Probe	28	**2329**	10	11	2378
U2R	10	9	**170**	8	197
R2L	113	49	29	**2535**	2726

Table 8. Confusion matrix obtained in the first phase over Kyoto 2006+ (Days, 2006, November 12-16)

		Predicted	
		Attack	Normal
Actual	Attack	20963	599
	Normal	50	6360

Similarly, the obtained results of Kyoto2006+ dataset are represented as confusion matrix in Table 9. Among the total 27972 test connections, the total number of connections classified as attacks are 20963 and the connections classified as normal are 6360 only. Based on the confusion matrices represented in Tables 4, 5 and 6, 7 and 8, the performance metrics are measured for both test sets and the obtained results are represented in Table 9.

Table 9. Performance analysis of proposed approach over KDDTest+ and Kyoto2006+ datasets

Metric	KDDTest+	KDDTest^{-21}	Kyoto2006+
DR (Recall)	97.7557	96.1749	99.2199
Precision	97.2655	97.3321	95.5998
Accuracy	98.9256	98.9745	97.9443
FAR	0.00458	0.0076	0.00780
F-Score	96.5025	97.0041	97.7876

Further the comparative analysis is carried out between the proposed and conventional approaches which followed the same dual methodology for the IDS. The comparison is done with respect to the DR, Precision, Accuracy, FAR and F-Score and the obtained values are represented in Table 10.

Table 10. Performance Comparison

Test set	Method	FAR (%)	Recall (%)	Precision (%)	F-Score (%)	Accuracy (%)
KDDTest+	Proposed	**0.0085**	**97.7557**	**97.2655**	**96.5025**	**98.9256**
	B.N.Kumar et.al [42]	0.0108	93.6622	99.1027	96.6893	98.7605
	SVM-ANN [33]	0.2135	89.4578	95.3028	91.9021	94.2335
	SVM-KPCA-GA [45]	0.2263	88.8563	93.2019	90.4389	92.2547
KDDTest^{-21}	Proposed	**0.0076**	**96.1749**	**97.3321**	**97.0041**	**98.9749**
	B.N.Kumar et.al [42]	0.0081	95.9123	99.3158	97.2356	98.8911
	SVM-ANN [33]	0.0107	91.1238	96.1888	93.7442	95.2335
	SVM-KPCA-GA [45]	0.0426	91.0217	94.5213	93.0106	93.8964
Kyoto2006+	Proposed	**0.0068**	**99.2199**	**95.5998**	**97.7879**	**97.9443**
	B.N.Kumar et.al [42]	0.0079	97.2232	94.3158	95.4764	97.2265
	SVM-ANN [33]	0.0144	91.1248	92.8884	91.2336	96.0012
	SVM-KPCA-GA [45]	0.0521	91.0523	92.1278	91.1787	94.3217

As it seen from Table 10, the DR of developed framework is more when compared it with the conventional approaches, signifying that the proposed mechanism detects more accurately. Further metrics also has favor towards the proposed approach. The recent method proposed by B.N.Kumaret.al., also accomplished hybrid intrusion detection mechanism by considering the SVM and CNN. Though these two classifiers has achieved a greater performance in the classification, the feature extraction technique (MI based FS) never considered the class irrelevancy. Due to this the features which are more relevant towards a particular class are removed. This problem is solved in the proposed approach and helps in the achievingthe optimal DR and classification accuracy. Further the conventional approaches such SVM-ANN [33], and SVM-KPCA-GA [45] are also hybrid techniques which tried to achieve an optimal performance by combining two algorithms. However, they are not focused on the FS technique by which an additional complexity arisen due to the increased number of features at the classifier. The proposed method also focused on this problem and developed a new FS algorithm by which the most relevant features are kept and the remaining features were removed. Due to the proposed FS mechanism, the detection performance at individual classes is increased and Figure 2 describes the details.

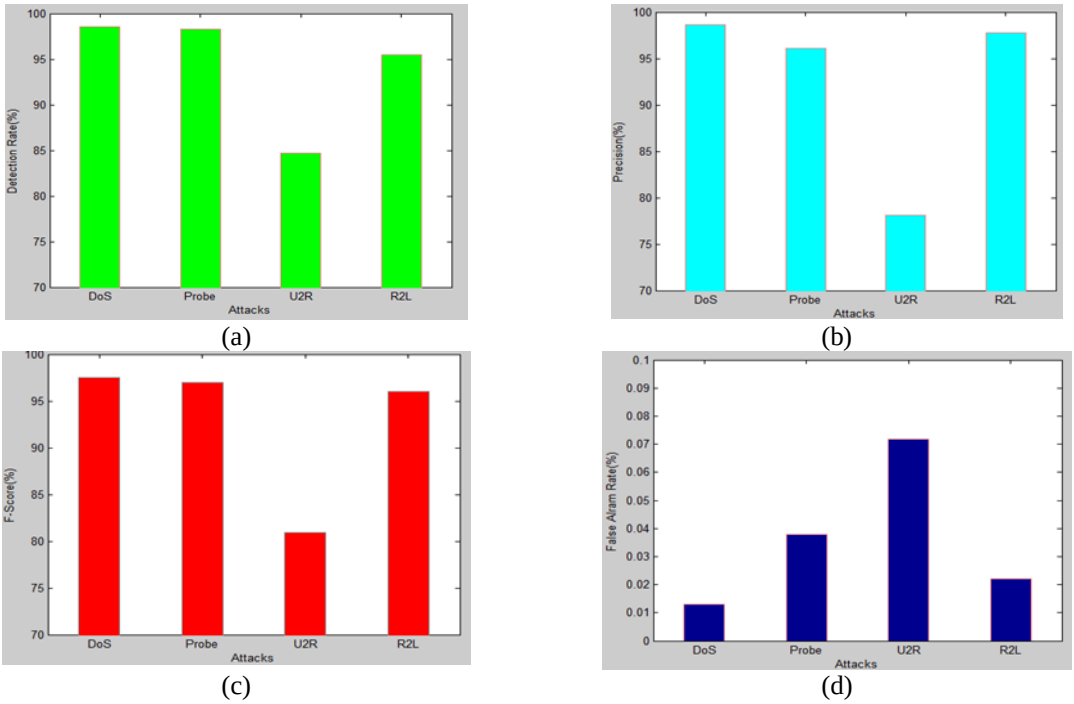

Figure 2. Performance analysis of Individual classes through propsoed approach by (a) DR, (b) Precision, (c) F-Score and (d) FAR

The obtained DR, precision, F-Score and FAR for individual classes is represented in Figure 2. Since the proposed system focused mostly on the feature sleection which provides a perfect discrmination between the individual classes, every testing connection can be classified more accurately by which the DR and precision increases more efficiently. Due to the cosnideration of class relevancy with every feature, the feature which were not significant with any information about the class is removed. Further the correct estimation of the significance of every feature is helped in reducing the FAR. As it is obsereved form Figure.2(d), the maximum FAR is 0.00728 and it is for rare attack, U2R. Compared to the conventional approaches, the FAR of proposed approach is seemed to be very less for both testsets.Hence the proposed approach is an optimal approach which provides an effective security for different applications.

Computational time is also an important aspect which needs to be a less value in the Intrusion Detection Strategy. As the technology increases, various types of attacks are increasing and to detect those attacks, the entire features needs to be trained to the system, which makes the system computationally time expensive. As the number of attacks followed by respective features is more, the time required to train and test the system increases. Here the proposed method selects only the required features by which most of the information signifies which makes the system computationally time inexpensive. The observed average timings for training testing of the proposed method are represented in Table 11. Table 11 also shows the comparison of times with for proposed and various conventional approaches. As it is observed from the table, the overall time of developed method is less compared to the conventional approaches. Though the proposed approach modeled LSTM-RNN classifier which consumes more time due to the feedback process at every state, the overall time is observed to be less due to the reduction of irrelevant features at the preprocessing.

Table 11. Average time for training and testing processes

Time (min)	Approach	Time
Training Time	SVM-ANN [33]	8.3325
	SVM-KPCA-GA [45]	9.9847
	B.N.Kumar et.al [42]	6.3347
	Proposed	5.4127
Testing Time	SVM-ANN [33]	6.1478
	SVM-KPCA-GA [45]	6.9898
	B.N.Kumar et.al [42]	4.3327
	Proposed	4.0023

6. CONCLUSION AND FUTURE SCOPE

Recent research on the IDSs has signified mainly two aspects which need to be achieved priory. They are (1) an efficient FS method and (2) a robust and simple method for classification. In this paper, a filter based FS algorithm (JMIMFS) combined with supervised learning is proposed. JMIMFS is an extended version of MIFS, MIFS-U and NMIFS. Compared to the conventional MI techniques, JMIMFS selects a more effective feature set which are more significant with every class and signify the most important information about every class. JMIMFS is then combined with the LSTM-RNN classifier to train the system. The LSTM-RNN is a deep learning technique which solves the non-linear dependencies between the variables. This process in carried out under the first phase and in the second phase, the Pearson correlation coefficient is accomplished as FS technique and the obtained features are trained through SVM algorithm. The extensive simulations carried out over the proposed approach through NSL-KDD and Kyoto2006+ datasets illustrates the effectiveness. The comparison between the earlier and proposed approaches reveals the enhancement in the detection performance. On an average the accuracy of proposed approach is improved by 3% in the NSL-KDD and 1.89% in the Kyoto2006+ dataset. Further, on an average, the computational time through developed framework is reduced by 3min when compared with conventional approaches.

Considering the deep characteristics of Kyoto2006+ dataset, further this work is extended to analyze different known and unknown attacks. In the Kyoto2006+ dataset, the unknown attacks also exist and most of the works did not focus in that direction. In the future, the further study of the Kyoto2006+ dataset will improvises the Intrusion detection at various levels of applications.

REFERENCES

[1] Y. Chen, A. Abraham, and B. Yang, "Hybrid flexible neural-tree based intrusion detection systems," *International Journal of Intelligent Systems*, vol. 22, no. 4, pp. 337–352, 2007.

[2] H.-J. Liao, C.-H. R. Lin, Y.-C. Lin, and K.-Y. Tung, "Intrusion detection system: A comprehensive review," *Journal of Network and Computer Applications*, vol. 36, no. 1, pp. 16 – 24, 2013.

[3] S.-Y. Wu and E. Yen, "Data mining-based intrusion detectors," *Expert Systems with Applications*, vol. 36, no. 3, Part 1, pp. 5605 – 5612, 2009.
[4] A. A. Aburomman and M. B. I. Reaz, "A survey of intrusion detection systems based on ensemble and hybrid classifiers," *Computers & Security*, vol. 65, pp. 135 – 152, 2017.
[5] O. Depren, M. Topallar, E. Anarim, and M. K. Ciliz, "An intelligent intrusion detection system (IDS) for anomaly and misuse detection in computer networks," *Expert Systems with Applications*, vol. 29, no. 4, pp. 713 – 722, 2005.
[6] Chandrashekar, G., &Sahin, F. (2014). A survey on feature selection methods. *Computers and Electrical Engineering, 40*, 16–28.
[7] Bolón-Canedo, V., Sánchez-Maroño, N., & Alonso-Betanzos, A. (2013). A review of feature selection methods on synthetic data. *Knowledge and Information Systems, 34*, 483–519.
[8] Lin, T., Li, H., & Tsai, K. (2004). Implementing the fisher's discriminant ratio in a k-means clustering algorithm for feature selection and data set trimming. *Journal of Chemical Information and Computer Sciences,44*,76–87.
[9] Battiti, R. (1994). Using mutual information for selecting features in supervised neuralnet learning. *IEEE Transactions on Neural Networks, 5*, 537–550.
[10] Liang, J., Wang, F., Dang, C., & Qian, Y. (2014). A group incremental approach to feature selection applying rough set technique. *IEEE Transactions on Knowledge and Data Engineering, 26* (2), 294–308.
[11] Zhang, Y., Yang, C.,Yang, A., Xiong, C.Y., Zhou, X., & Zhang, Z. (2015). Feature selection for classification with class-separability strategy and data envelopment analysis. *Neurocomputing, 166*, 172–184.
[12] Beauquier, J. and Hu, Y. (2007) 'Intrusion detection based on distance combination', *CESSE07*, Venice, Italy, World Academy of Sciences, WAS.
[13] Jin, S., Yeung, D.S. and Wang, X. (2007) 'Network intrusion detection in covariance feature space', *Pattern Recognition*, Vol. 40, No. 2, pp.2185–2197.
[14] Akshadeep, Ishfaq Manzoor, Neeraj Kumar, "A feature reduced intrusion detection system using ANN classifier", Expert systems with applications, Vol.88, Dec 1, 2017, pp.249-257.
[15] ChaoukiKhammassi and Saoussen, "A GA-LR wrapper approach for feature selection in network intrusion detection", Computer Security, Vol.70, September 2017, pp.255-277.
[16] Zhu, Yingying, Liang, Junwei, Chen, Jian, yong, Zhong, Ming, "An improved NSGA-III algorithm for feature selection used in intrusion detection", Knowledge Based systems, Vol.116, January 15, 2017, pp.74-85.
[17] F. Amiri, M. RezaeiYousefi, C. Lucas, A. Shakery, N. Yazdani, Mutual information-based feature selection for intrusion detection systems, Journal of Network and Computer Applications 34 (4) (2011) 1184–1199.
[18] SumaiyaThaseen and C. A. Kumar, "Intrusion Detection Model using fusion of PCA and optimized SVM", *In: Proc. of International Conf. on Computing and Informatics (IC3I)*, Mysore, India, 2014, pp. 879–884.
[19] SumiyaThaseen ikram and C. A. Kumar, "Intrusion Detection Model using fusion of chi-square feature selection and multi class SVM", *Journal of King Saud University – Computer and Information Sciences*, Vol.29, 2017, pp.462-472.
[20] SumaiyaThaseen ikram and C. A. Kumar, "Intrusion Detection model using chi-square feature selection and modified naïve Bayesian classifier", In: *Proc. of third International Symposium on Big Data and Cloud Computing challenges*, 2016, pp.81-91.
[21] Harshit Saxena, Vineet Richariya, "Intrusion Detection in KDD99 dataset using SVM-PSO and feature reduction with information gain", Int. J. COmput. Appl., 98 (6), 2014, pp.25-29.
[22] Mohammed A. Ambusaidi, Xiangjian He, Priyadarsi Nanda, and Zhiyuan Tan, "Building an intrusion detection system using a filter-based feature selection algorithm", IEEE Transactions on Computers, November 2014.
[23] A. A. Aburomman and M. B. I. Reaz, "A novel SVM-kNN-PSO ensemblemethod for intrusion detection system," *Applied Soft Computing*,vol. 38, pp. 360–372, 2016.
[24] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A fast and elitistmultiobjective genetic algorithm: NSGA-II," *IEEE Transactions onEvolutionary Computation*, vol. 6, no. 2, pp. 182–197, 2002.
[25] S.-Y. Ji, B.-K. Jeong, S. Choi, and D. H. Jeong, "A multi-level intrusiondetection method for abnormal network behaviors," *Journal of Networkand Computer Applications*, vol. 62, pp. 9 – 17, 2016.
[26] S. M. H. Bamakan, H. Wang, T. Yingjie, and Y. Shi, "An effectiveintrusion detection framework based on MCLP/SVM optimized by timevaryingchaos particle swarm optimization," *Neurocomputing*, vol. 199,pp. 90–102, 2016.
[27] W. L. Al-Yaseen, Z. A. Othman, and M. Z. A. Nazri, "Multi-levelhybrid support vector machine and extreme learning machine based onmodified k-means for intrusion detection system," *Expert Systems withApplications*, vol. 67, pp. 296 – 303, 2017.
[28] J. R. Quinlan, "Introduction of decision trees," *Machine Learning*, vol. 1,no. 1, pp. 81–106, 1986.
[29] R. A. R. Ashfaq, X.-Z. Wang, J. Z. Huang, H. Abbas, and Y.-L.He, "Fuzziness based semi-supervised learning approach for intrusiondetection system," *Information Sciences*, vol. 378, pp. 484 – 497, 2017.
[30] H. Bostani and M. Sheikhan, "Modification of supervised OPF-based intrusion detection systems using unsupervised learning and social network concept," *Pattern Recognition*, vol. 62, pp. 56–72, 2017.
[31] S.-W. Lin, K.-C. Ying, C.-Y. Lee, and Z.-J. Lee, "An intelligent algorithm with feature selection and decision rules applied to anomaly intrusion detection," *Applied Soft Computing*, vol. 12, no. 10, pp. 3285– 3290, 2012.
[32] W.-C. Lin, S.-W. Ke, and C.-F. Tsai, "CANN: An intrusion detection system based on combining cluster centers and nearest neighbors," *Knowledge-Based Systems*, vol. 78, pp. 13 – 21, 2015.
[33] Samuel Lalmuanawma, "A two-stage hybrid classification technique fornetwork intrusion detection system", International Journal of Computational Intelligence Systems, 9:5, pp. 863-875, 2016 .
[34] Hochreiter, Sepp, and Jrgen Schmidhuber, *Long short-term memory*, Neural computation 9.8, pp.1735-1780, 1997

[35] LukoEviIus, Mantas, and Herbert Jaeger, *Reservoir computing approachesto recurrent neural network training*, Computer Science Review3.3 pp.127-149, 2009
[36] Kwok,N., &Choi,C. (2002). Input feature selection for classification problems. *IEEE Transactionson Neural Networks, 13*, 143–159.
[37] Peng,H., Long,F., & Ding,C. (2005). Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Transactionson Pattern Analysis and Machine Intelligence, 27*, 1226–1238.
[38] Estévez,P.A., Tesmer,M., Perez,A., &Zurada,J.M. (2009). Normalized mutual information feature selection. *IEEE Transactions on Neural Networks, 20*, 189–201.
[39] Hoque,N., Bhattacharyya, D.K., & Kalita,J.K. (2014). MIFS-ND: a mutual information-based feature selection method. *Expert Systems with Applications, 41* (14), 6371–6385.
[40] Yang,H., & Moody,J. (1999). Feature selection based on joint mutual information. In *Proceedings of international ICSC symposium on advances in intelligent data analysis* (pp.22–25).
[41] Bukka Narendra Kumar, Mantena S. V. Sivarama Bhadri Raju, Bulusu Vishnu Vardhan, "Enhancing the Performance of an Intrusion Detection System Through Multi- Linear Dimensionality Reduction and Multi-Class SVM", *International Journal of Intelligent Engineering and Systems, Vol.11, No.1, 2018.*
[42] Bukka Narendra Kumar, Mantena S. V. Sivarama Bhadri Raju, Bulusu Vishnu Vardhan, "Selective Feature Mechanism for Two-Phase Intrusion Detection System", Jour of Adv Research in Dynamical & Control Systems, Vol. 10, 02-Special Issue, 2018
[43] M. Tavallaee, E. Bagheri, W. Lu, and A.-A. Ghorbani, "A detailed analysis of the KDD CUP 99 data set," in Proceedings of the Second IEEE Symposium on Computational Intelligence for Security and Defense Applications 2009, 2009.
[44] J. Song, H. Takakura, Y. Okabe, M. Eto, D. Inoue, K. Nakao, Statistical analysis of honeypot data and building of kyoto 2006+ dataset for nids evaluation, in: Proceedings of the First Workshop on Building Analysis Datasets and Gathering Experience Returns for Security, ACM, 2011, pp. 29–36.
[45] F. Kuang, W. Xu, and S. Zhang, "A novel hybrid KPCA and SVM with GA model for intrusion detection," *Applied Soft Computing*, vol. 18, pp. 178 – 184, 2014.