Improving the Performance of Network Intrusion Detection Based on Hybrid Feature Selection Model

Joseph MBUGUA[1], Joseph SIROR[2], Moses THIGA[3]

Kabarak University, 13 P.O. Box Private Bag, Kabarak, 20157, Kenya
[1]Jmbugua80@yahoo.com,[2]josephsiror@gmail.com[3] mthiga@kabarak.ac.ke

Abstract:
Due to the high dimensionality of the network traffic data, it is not realistic for an Intrusion Detection System (IDS) to detect intrusions quickly and accurately. Feature selection is an essential component in designing intrusion detection system to eliminate the associated shortcoming and enhance its performance through the reduction of its complexity and acceleration of the detection process. It eliminates irrelevant and repetitive features from the dataset to make robust, efficient, accurate and lightweight intrusion detection system to be certain timelines for real time.In this paper, a novel feature selection model is proposed based on hybridising feature selection techniques (information gain, correlation feature selection and chi square). In the experiment the performance of the proposed feature selection model is tested with different evaluation metrics which includes: True Positive rate (TR), Precision (Pr), false positive rate (FPR), on NSL KDD datasetwith four different classificaton techniques i.e. random forest, Bayes, J48, Parts. The experimental results showed that the proposed model improves the detection rates and also speed up the detection process.
Key words: Intrusion detection, Performance, hybrid, feature selection, classifier.

Introduction

Network Intrusion Detection System (IDS) [1]monitors the use of computers and networks over which they communicate , searching for unauthorised use, anamolous behaviour, and attempt to deny users, machines or portions of networks access to the services. Although the intrusion detection systems are increasingly deployed in the computer network, they deal with a huge amount of data that contains null values, incomplete information, and irrelevant features. The analysis of the large quantities of data can be tedious, time-consuming and error-prone. Data mining and machine learning[2] provides tools to select best relevance features subset which improves detection accuracy and removes distractions.
Feature selection procedures require four basic stages in a simple feature selection method [3].
(1) Generation procedure in order to generate the upcoming candidate subset
(2) Evaluation function so that it can evaluate the subset
(3) Stopping criterion to decide when to stop
(4) Validation procedure used for validates the subset.

The existing feature selection techniques in machine learning can be broadly cassified into two categories i.e wrappers and filters. Wrappers selection techniques evaluate the worth of features using the learning algorithm applied to the data while filters evaluate the worth of features by using heuristics based on general characteristics of the data. Feature selection algorithms can be further differentiated by the exact nature of their evaluation function, and by how the space of feature subsets is explored. Wrappers often give better results in terms of the final predictive

accuracy of a learning algorithm than filters because feature selection is optimized for the particular learning algorithm used. However, since a learning algorithm is employed to evaluate each and every set of features considered, wrappers are prohibitively expensive to run, and can be intractable for large databases containing many features. Furthermore, since the feature selection process is tightly coupled with a learning algorithm, wrappers are less general than filters and must be re-run when switching from one learning algorithm to another.

The advantages of filter approaches in feature selection outweigh their disadvantages. Flters execute many times faster as compared to wrappers and therefore applicable in databases with a large number of features [4]. They do not require re-execution for different learning algorithms and can provide an intelligent starting feature subset for a wrapper in case improved accuracy for a particular learning algorithm is required[5]. Filter algorithms also exhibited a number of drawbacks. Some algorithms do not handle noise in data, and others require that the level of noise be roughly specified by the user a-priori [5], [6]. In some cases, a subset of features is not selected explicitly; instead, features are ranked with the final choice left to the user. In other cases, the user must specify how many features are required, or must manually set a threshold by which feature selection terminates. Some algorithms require data to be transformed in a way that actually increases the initial number of features. This last case can result in a dramatic increase in the size of the search space [6].

The rest of the paper is organized as follows: Section II presents some related researches on intrusion detection which cover the feature selection and data mining. Section III briefly describes the KDD dataset used in this research. Section IV explains the details of the dataset pre-processing phase of the proposed model. The proposed model is presented in Section V. Finally, the experimental results and analysis are presented in Section 6 followed by some conclusions in the final section.

RELATED WORK

Recent study indicates that machine learning algorithms can be adversely affected by irrelevant and redundant training information [7]. The simple nearest neighbour algorithm is sensitive to irrelevant attributes, its sample complexity (number of training examples needed to reach a given accuracy level) grows exponentially with the number of irrelevant attributes[8][9]. Sample complexity for decision tree algorithms can grow exponentially on some concepts (such as parity) as well. The naive Bayes classifier can be adversely affected by redundant attributes due to its assumption that attributes are independent given the class [10] . Decision tree [11], [12] algorithms such as C4.5 overfit training data, resulting in large trees. In many cases, removing irrelevant and redundant information can result in C4.5 producing smaller trees. As a result, most researchers combines the feature selection and classification algorithms to improve the detection accuracy and make intelligent decisions in determining intrusions. Siraj et al. [16] proposed new, automated and intelligent hybrid clustering model called Improved Unit Range and Principal Component Analysis with Expectation Maximization (IPCA-EM) to aggregate similar alerts as well as to filter the low quality alerts. Panda et al. [2] proposed a hybrid intelligent approach using combination of classifiers in order to make the decision intelligently, so that the overall performance of the resultant model is enhanced. These two models use hybrid classifiers to make intelligent decisions and the filtering process is applied after adding

supervised or unsupervised learning techniques to obtain the final decision. Agarwal et al. [47] proposed hybrid approach for anomaly intrusion detection system based on combination of both entropy of important network features and support vector machine.

Lin et al. (2015) studied the importance of feature representation method on classification process. They proposed cluster centre and nearest neighbour (CANN) approach as a novel feature representation approach. In their approach, they measured and summed two distances. The first distance measured the distance between each data sample and its cluster centre. The second distance measured the distance between the data and its nearest neighbour in the same cluster. They used this new one-dimensional distance to represent each data sample for intrusion detection by a k-nearest neighbour (k-NN) classifier. The proposed approach provided high performance in terms of classification accuracy, detection rates, and false alarms. In addition, it provided high computational efficiency for the time of classifier training and testing

Methodology

The proposed model has four phases as shown in figure 1:
Phase 1 data pre-processing
Phase 2 feature selection techniques.
Phase 3 classification techniques.
Phase 4 evaluation.

Data Preprocessing

To make efficient use of the available dataset for analysis the data preprocessing is required to provide solutions to Clean the data to remove noise and duplicate information and then deal with any incomplete or missing data an efficient algorithm based on normalization and discretization techniques. Data normalizaion is a process of scaling the value of each feature into a well-proportioned range, so that the bias in favor of features with greater values is eliminated from the dataset [13]. Every attribute within each record is scaled by the respective maximum value and falls into the same range of [0-1]. Normalization follows equation 1,

$$\text{Normalized}(x_i) = \frac{(X_i - X_{min})}{(X_{Max} - X_{min})} \quad \ldots\ldots\ldots\ldots\ldots\ldots\text{Eq. (1)}$$

where X_{min} is the minimum value for variable X, X_{max} is the maximum value for variable X. For a specific symbolic feature, we assigned a discrete integer to each value and then used equation 1 to normalize it.

Discretization transforms continuous valued attributes to nominal [14][15]. The main benefit is that some classifiers can only take nominal attributes as input, not numeric attributes and also some classifiers can only take numeric attributes and hence can achieve improved accuracy if the data is discretized prior to learning.

Feature Selection Techniques

The feature selection techniques help to identify some of the important attributes in a data set, thus reducing the memory requirement, increase the speed of execution and improves the classification accuracy[16]. The purpose of this work is to find out which data feature selection algorithm gives better results with decision trees classifiers. Several feature subset selection techniques have been used in data mining.

i. Correlation based feature selection (CFS)

CFS is considered as one of the simplest yet effective feature selection method which is based on the assumption that features are conditionally independent given the class, where feature subsets are evaluated according to a correlation based heuristic evaluation function.[17]. A good feature subset is one that contains features highly correlated with the class, yet uncorrelated with each other. The major advantage of CFS, it is a filter algorithm, which makes it much faster compared to a wrapper selection method since it does not need to invoke the learning algorithms [18],[19].

$$\rho(X,Y) = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\left[\sum_i (x_i - \bar{x})^2 \sum_i (y_i - \bar{y})^2\right]^{\frac{1}{2}}} \quad \text{.....(2)}$$

Pearson's correlation coefficient (2), where all variables have been standardized shows that the correlation between a composite and an outside variable is a function of the number of component variables in the composite and the magnitude of the inter-correlations among them, together with the magnitude of the correlations between the components and the outside variable.

ii. Information Gain

Information gain is used as a measure for evaluating the worth of an attribute based on the concept of entropy (1), the higher the entropy the more the information content. Entropy can be viewed as a measure of uncertainty of the system. The largest mutual information between each feature and a class label within a certain group is then selected (2). The performance evaluation results show that better classification performance can be attained from such selected features [20],[18].

$$-\sum_i P(c_i)\log_2 P(c_i). \qquad \text{Eq. (3)}$$

$$IG(A) = I(D) - \sum_{j=1}^{p} \frac{|D_j|}{|D|} I(D_j^A) \qquad \text{Eq. (4)}$$

Algorithm 1: Feature selection according to information gains

Input: A training dataset T = D(F,C), number of features to be selected L

Output: Selected features S

1. Initialize relative parameters: F← fi, i =1, 2, ...n, C←'class labels', S =? ;
2. for each feature fi ∈ F do
   a. Calculate its information gain IG(fi) ;
   b. insert fi into S in descending order with regard to IG(fi) ;
3. Retain first L feature in S, and delete the others ;
4. Return Selected features: S.

iii. Chi-square

Chi-square [18] test is commonly used method, which evaluates features individually by measuring chi- square statistic with respect to the classes. The statistic is

$$\sum_{i=1}^{k} \sum_{i=1}^{k} \frac{(A_{ij} - E_{ij})^2}{E_{ij}}$$

value i for attribute and j for the class,

Eij = the expected No. of instances for Aij.

The larger value of the χ^2, indicates highly predictive to the class.

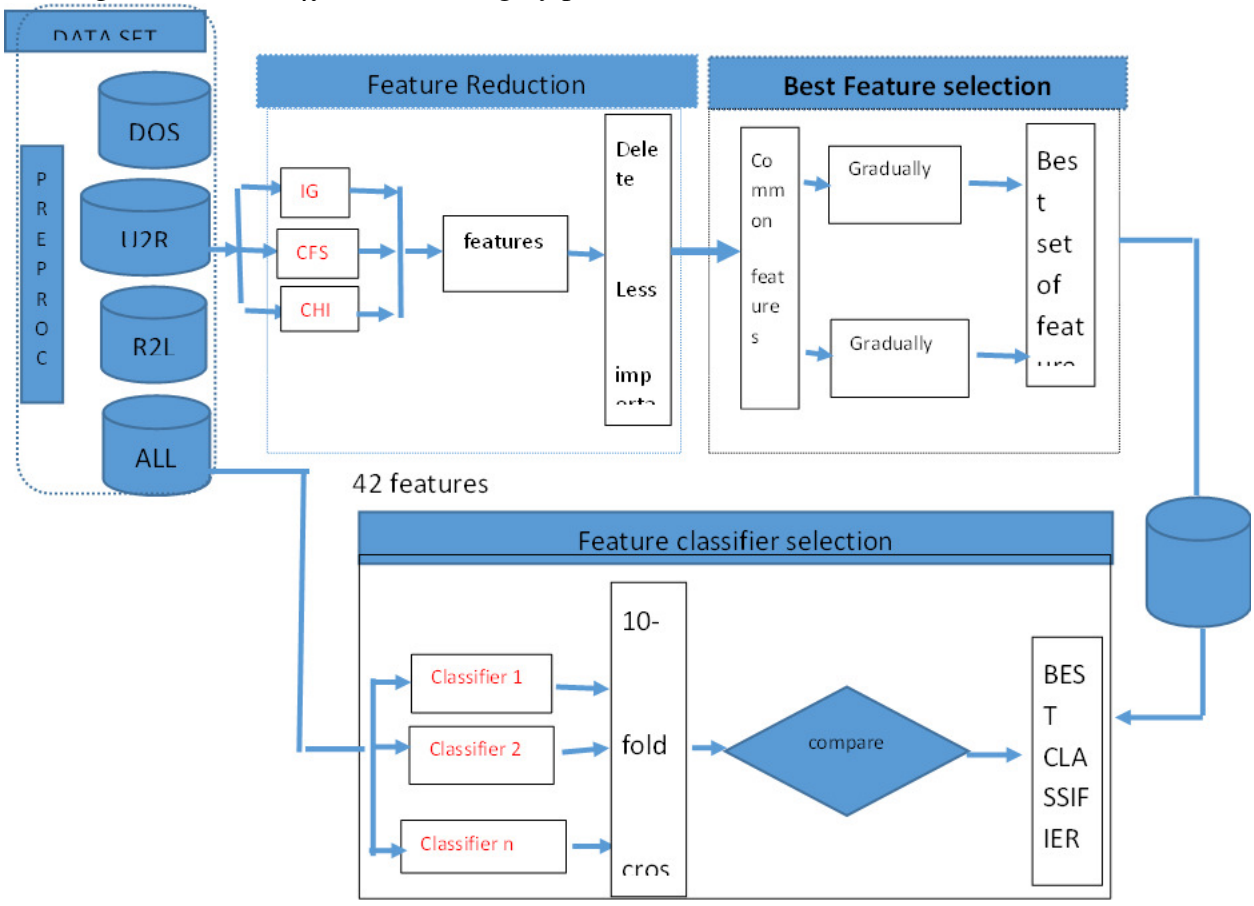

Figure 1 best feature selection process

Classification techniques

Classification [21] is a machine learning technique where similar type of samples are grouped together in supervised manner and can classify the intrusion data as normal or attack.

Random Forest

Random Forest is an ensemble learning technique for classification and predictive modeling. It is also an approach to data exploration and generates many trees by using recursive partitioning

then aggregate the results[22]. Each of the trees is constructed separately by using a bootstrap sample of the data and the bagging technique[23] is applied to combine all results from each of the trees in the forest. The method used to combine the results can be as simple as predicting the class obtained from the highest number of trees.

Bayesian Network

Bayesian reasoning provides a probabilistic approach for inference and is based on the assumption that the quantities of interest are governed by probability distributions and that optimal decisions can be made by reasoning about these probabilities together with observed data [24]. A Bayesian network is a graphical model that encodes probabilistic relationships among variables of interest. When used in conjunction with statistical techniques, bayesian networks have several advantages for data analysis (Kaur & Sachdeva, 2016;Assi & Sadiq, 2017). First, the Bayesian networks encode the interdependencies between variables and hence they can handle situations where data are missing. Secondly, the Bayesian networks have the ability to represent causal relationships. Therefore, they can be used to predict the consequences of an action. Lastly, the Bayesian networks have both causal and probabilistic relationships; they can be used to model problems where there is a need to combine prior knowledge with data. The disadvantages of Bayesian networks includes [27]. First, the classification capability of naïve Bayesian networks is identical to a threshold-based system that computes the sum of the outputs obtained from the child nodes. Secondly, the child nodes do not interact between themselves and their output only influences the probability of the root node and hence incorporating additional information becomes difficult as the variables that contain the information cannot directly interact with the child nodes. Lastly, the accuracy of this method is dependent on certain assumptions that are typically based on the behavioral model of the target system and deviating from those assumptions will decrease its accuracy. Therefore, selecting an accurate model will lead to an inaccurate detection system as typical systems and/or networks are complex.

J48

J48 [22] is an open source Java implementation of the C4.5 algorithm in the WEKA data mining tool. C4.5 is a program that creates a decision tree based on a set of labeled input data. The decision trees generated by C4.5 can be used for classification, and for this reason, C4.5 is often referred to as a statistical classifier. J48 classifier algorithms [26] are used to compare and built, using the information entropy process, a decision tree from a set of training dataset. These algorithms adopt a top down technique and inductively built the decision tree for classification. It's extremely efficient when handling large datasets. [28]. The extra features of J48 [29] includes accounting for missing values, decision trees pruning, continuous attribute value ranges and derivation of rules.

To make actual decisions regarding which path of the tree to replace is based on the error rates used. The reserved portion can be used as test data for the decision tree to overcome potential overfitting problem (reduced-error pruning).

Now, among the possible values of this feature, if there is any value for which there is no ambiguity that is for which the data instances falling within its category have the same value for the target variable then terminate that branch and allocate to it the target value that have obtained.

EXPERIMENT & DISCUSSION

Data Set

The experiments is conducted on MIT Lincoln's Lab's DARPA 2000 Scenario Specific NSL-KDD, 2014 which contains simulated attack scenarios in a protected environment an off-site server. KDD"99 testing set includes 37 attack types that are included in the testing set.

The simulated attacks in the NSL-KDD dataset fall in one of the following four categories[8], [30]–[32].

i. Denial of service attack (Dos), where attempts are to shut down, suspend services of a network resource remotely making it unavailable to its intended users by overloading the server with too many requests to be handled. e.g. syn flooding. Relevant features includes source bytes and percentage of packets with errors. Examples of attacks includes back,land, neptune, pod, smurf, teardrop

ii. Probe attacks, where the hacker scans the network of computers or DNS server to find valid IP, active ports, host operating system and known vulnerabilities with the aim discover useful information. Relevant features includes duration of connection and source bytes. Examples includes Ipsweep, nmap, portsweep, satan

iii. Remote-to-Local (R2L) attacks, where an attacker who does not have an account with the machine tries to gain local access to unauthorized information through sending packets to the victim machine exfiltrates files from the machine or modifies in transist to the machine. Relevant features includes number of file creations and number of shell prompts invoked. Attacks in this category includes ftp_write, guess_passwd, imap, multihop, phf, spy, warezclient, warezmaster

iv. User-to-Root (U2R) attacks, where an attacker gains root access to the system using his normal user account to exploit vulnerabilities. Relevant features includes Network level features – duration of connection and service requested and host level features - number of failed login attempts. Attacks includes buffer_overflow, loadmodule, perl, rootkit

Experimental Setup

In the experiment, we apply full dataset as training set and 10-fold cross validation for the testing purposes. The available dataset is randomly subdivided into 10 equal disjoint subsets and one of them is used as the test set and the remaining sets are used for building the classifier. In this process, the test subset is used to calculate the output accuracy while theN_1 subset is used as a test subset and to find the accuracy for each subset. The process is repeated until each subset is used as test set once and to compute the output accuracy of each subset. The final accuracy of the system is computed based on the accuracy of the entire 10 disjoint subsets.

For our experiment, we selected attribute set based on the repetition of attribute from four scheme. Existing FS that are employed in experiments are Correlation Feature Selection (CFS) based evaluator with Best-first searching method, Gain Ratio (GR) Attributes based Evaluator with Ranker searching method, Information Gain (IG) based Attributes Evaluator with ranker searching method, and Chi Squared Eval and Ranker searching method we obtained.

Table 1: The best set of relevant features

ALL	8	2,3,4,26,27,36,39,41
DOS	9	2,3,9,26,41,4,26,27,41
PROBE	7	2,3,9,30,34,38,40
R2L	8	1,2,7,33,3,40,34,30,21
U2R	7	6,11,29,30,3,10,14
Best Feature	1 2	1,2,3,9,26,27,29,30,34,36,3 9,40

Evaluation Metrics

The performance of the proposed feature selection technique is tested with different evaluation metrics such as: True Positive rate (TR), accuracy (Acc) also known as correctly classified instances (CC), Precision (Pr), false alarm rate (FAR) also known as false positive rate (FPR), ROC and F measure rate (MR).

i. True positive rate (TPR):

TP/(TP+FN), also known as detection rate (DR) or sensitivity or recall.

ii. False positive rate (FPR):

FP/(TN+FP) also known as the false alarm rate.

iii. Precision (P):

TP/(TP+FP) is defined as the proportion of the true positives against all the positive results.

iv. Total Accuracy (TA):

(TP+TN)/(TP+TN+FP+FN) is the proportion of true results (both true positives and true negatives) in the population.

v. F-measure: 2PR/(P+R) is the harmonic mean of precision and recall

Experimental Results and Analysis

In this experiment the time required for the classifiers to build the training model based on several feature selection techniques, namely: ALL, CFS, chi-squared, Information Gain and proposed features are compared with four different classifiers i.e. random forest, Bayes, J48, Parts as shown in Table 8. The experiments indicate that using random forest as a classifier in the training phase takes longer time train the model and hence the results can best illustrate the enhancement of the proposed feature selection technique in the overall performance of intrusion detection. Using all 41 features without selecting important features increases the overhead of the classifiers which subsequently increases the time to build the model. It can be observed that correlation features selection has the most efficient time across most classifiers with exception of chi-squared and Bayes Net (0/16). J48, Bayes and parts presents the most time efficient classification algorithm for all the filter selection methods. As observed, the proposed scheme outperforms the existing techniques with significantly less training time with exception of correlation based feature selection technique and the performance of the overall system is not degraded and its effectiveness is not compromised.

Table 8 Training times for different feature selection techniques

Feature selection	Bayes	J48	Random forest	PART
CFS	0.21	0.32	5.68	1.38
ALL	0.8	2.16	10.04	2.39
CHI	0.16	0.95	9.86	5.24
IG	0.28	0.98	9.52	5.07
PROPOSED	**0.35**	**0.82**	**6.99**	**1.36**

The work also compares the performance of proposed technique in terms of True Positive rate, false positive rate, precision, F measures and ROC with other schemes as indicated in Table 7. Comparing our proposed technique against using the full dataset with 42 features, the table indicates some enhancement has been obtained and even no degradation is observed. For instance, the false positive and accuracy have been decreased around 3% and improved about 5% respectively. Additionally, it is shown that the proposed technique has the best performance among other feature selection techniques.

Table 10 and Table 11 presents the performance of different classification algorithms with the proposed selected feature and full dataset, respectively. It is apparent that, with regard to the decrease in the training time, the overall performance of the model is enhanced comparing to using 42 features although with some exceptions. The false alarm rate for PART classifier, for example, with an increase about 1 % has a negative impact on the performance of proposed features set, however since the other evaluation metrics has increased or maintained at the same level, this adverse impact can be considered negligible. Moreover, among all classifiers, J48 classifier has the highest accuracy and precision and the lowest miss and false alarm rate, hence considered to be the best classifier for the proposed feature selection technique.

Table 11 and 12 demonstrate the performance of different feature selection techniques tested with different classification algorithms namely: J48, Random forest, PART and Bayesian in terms of detection rate and false positive rate, respectively. As it can be observed, regardless of the classification algorithms, the performance of the proposed technique significantly improves comparing to other schemes. For instance, the detection rate of our proposed technique is averagely 99% for all the classifiers as compared to the ALL, CFS, CHI and IG feature selection technique with average detection rate of 98 %, 98%, 91% AND 92% respectively. Similarly, Table 9 shows that the false alarm rate for the proposed technique with a significant drop comparing to other schemes helps to enhance the performance. For example, on average the proposed scheme with 1% has significantly less false alarm rate than ALL, CFS, CHI and IG feature selection technique with 2, 2, 10 and 8 %, respectively.

Classification Results Using All 42 Features of NSL-KDD Dataset

Table 9 Classification Results Using All Features of NSL KDD Dataset

Classifier	TP	FP	PR	RECALL	FM	ROC
J48	99	0.8	98	97	97	99
R Forest	99	0.8	96	98	98	99
Bayesian	97	1.7	97	97	97	99
PART	99	0.7	98	98	97	99

Table 10 Classification Results Using Proposed Features of NSL KDD Dataset

Classifier	TP	FR	PR	RECALL	FM	ROC
J48	99	0.4	99	99	99	99
R Forest	99	0.2	99	99	99	1
Bayesian	95	1.6	96	95	96	99
PART	99	0.3	99	99	99	99

Table 11 The Performance of five Feature Selection Techniques with Different Classifier in Terms of Detection Rate

FST	J48	RF	PART	BAYES
Full	99	99	99	95
CFS	99	99	98	97
Chi square	92	93	93	88
IG	93	93	93	88
PROPOSED	99	99	99	97

Table 12 The performance of feature selection techniques with different classifier in terms of false positive rates

FST	J48	RF	PART	BAYESIAN
FULL	1	1	1	5
CFS	1	1	2	3
PROPOSED	1	1	1	3
CHI	8	7	7	12
IG	7	7	7	12

Conclusions

This work examines the features included in NSL- KDD dataset to identify the significant features and reduce the number of features in the NSL- KDD dataset. Therefore, a subset of significant features in detecting intrusion can be proposed by using machine learning techniques. These features then can be used in the design of Intrusion Detection Systems (IDS), working towards automating anomaly detection with less overhead. The most important features to detect cyber-attacks are basic features such as source byte, destination byte, the used service, a flag to indicate the status of the connection. Moreover, time-based traffic features are important to analysis and detect cyber-attacks, such as information about the percentage of connections in the past 2 seconds with a different service than current connection. To detect R2L and U2R attacks it is important to study content features.

The proposed feature selection technique is compared with other well-known feature selection algorithms namely: CFS, IG, and chi-squared on NSL-KDD dataset. The results indicate that the proposed technique has considerably less training time while maintaining accuracy and precision. In addition, to demonstrate the effect of pre-processing dataset on classification rate using filter feature selection methods, different feature selection techniques are tested with four different classification algorithms namely: J48, Random forest, PART and Bayesianin terms of detection rate and False Alarm Rate. Regardless of the classification algorithm, the results indicate that the proposed scheme out performs other techniques. Another observation from comparison results between the proposed technique and using the full dataset is that J48 classification algorithm performs better with proposed feature selection algorithm than other classifiers. This is expected as random forest is an ensemble classifier that combines a collection of classifiers to make a forest.

References

[1] M. M. Siraj, H. Hussein, T. Albasheer, and M. M. Din, "Towards Predictive Real-time Multi-sensors Intrusion Alert Correlation Framework," *Indian J. Sci. Technol. ISS*N, vol. 8, no. 12, pp. 974–6846, 2015.

[2] M. C. Belavagi and B. Muniyal, "Performance Evaluation of Supervised Machine Learning Algorithms for Intrusion Detection," *Procedia Comput. Sci.*, vol. 89, pp. 117–123, 2016.

[3] H. Singh and D. Kumar, "A study on Performance analysis of various feature selection techniques in intrusion detection system," *Int. J. Adv. Res. Comput. Sci. Manag. Stud.*, vol. 3, no. 6, pp. 2321–7782, 2015.

[4] M. Othman and T. Maklumat, "Mobile Computing and Communications: An Introduction," *Malaysian J. Comput.* …, vol. 12, no. 2, pp. 71–78, 1999.

[5] K. Kumar, "Network Intrusion Detection with Feature Selection Techniques using Machine-Learning Algorithms," vol. 150, no. 12, pp. 1–13, 2016.

[6] J. Song, "Feature Selection for Intrusion Detection System Jingping Song Declaration and Statement," p. 132, 2016.

[7] N. A. Noureldien and I. M. Yousif, "Accuracy of Machine Learning Algorithms in Detecting DoS Attacks Types," vol. 6, no. 4, pp. 89–92, 2016.

[8] A. Thesis, "Using Support Vector Machines in Anomaly Intrusion Detection by," 2015.

[9] P. Verma, "Performance of Detection Attack using IDS Technique," vol. 4, no. 3, pp. 624–629, 2016.

[10] J. Juanchaiyaphum, N. Arch-int, and S. Arch-int, "A Novel Lightweight Hybrid Intrusion Detection Method Using a Combination of Data Mining Techniques," *Int. J. Secur. its Appl.*, vol. 9, no. 4, pp. 91–106, 2015.

[11] P. Manandhar, "A Practical Approach to Anomaly - based Intrusion Detection System by Outlier Mining in Network Traffic By," 2014.

[12] A. I. Madbouly, A. M. Gody, and T. M. Barakat, "Relevant Feature Selection Model Using Data Mining for Intrusion Detection System," *Int. J. Eng. Trends Technol.*, vol. 9, no. 10, pp. 501–512, 2014.

[13] M. A. Ambusaidi, X. He, Z. Tan, P. Nanda, L. F. Lu, and U. T. Nagar, "A Novel Feature Selection Approach for Intrusion Detection Data Classification," *2014 IEEE 13th Int. Conf. Trust. Secur. Priv. Comput. Commun.*, pp. 82–89, 2014.

[14] D. a. M. S. Revathi, "A Detailed Analysis on NSL-KDD Dataset Using Various Machine Learning Techniques for Intrusion Detection," *Int. J. Eng. Res. Technol.*, vol. 2, no. 12, pp. 1848–1853, 2013.

[15] S. K. Sahu, S. Sarangi, and S. K. Jena, "A detail analysis on intrusion detection datasets," *Souvenir 2014 IEEE Int. Adv. Comput. Conf. IACC 2014*, no. December, pp. 1348–1353, 2014.

[16] Z. Dewa and L. A. Maglaras, "Data Mining and Intrusion Detection Systems," *Int. J. Adv. Comput. Sci. Appl.*, vol. 1, no. 1, p. 1:7, 2016.

[17] Y. Wahba, E. ElSalamouny, and G. ElTaweel, "Improving the Performance of Multi-class Intrusion Detection Systems using Feature Reduction," *Ijcsi*, vol. 12, no. 3, pp. 255–262, 2015.

[18] V. Barot, S. Singh Chauhan, and B. Patel, "Feature Selection for Modeling Intrusion Detection," *Int. J. Comput. Netw. Inf. Secur.*, vol. 6, no. 7, pp. 56–62, 2014.

[19] M. B. Shahbaz, X. Wang, A. Behnad, and J. Samarabandu, "On Efficiency Enhancement of the Correlation-based Feature Selection for Intrusion Detection Systems," 2016.

[20] A. AliShah, M. Sikander Hayat Khiyal, and M. Daud Awan, "Analysis of Machine Learning Techniques for Intrusion Detection System: A Review," *Int. J. Comput. Appl.*, vol. 119, no. 3, pp. 19–29, 2015.

[21] S. Thaseen and C. A. Kumar, "Intrusion Detection Model using PCA and Ensemble of Classifiers," vol. 16, no. 2, pp. 15–38, 2016.

[22] R. Pradhan, "Performance Assessment of Robust Ensemble Model for Intrusion Detection using Decision Tree Techniques," vol. 3, no. 3, pp. 78–86, 2014.

[23] S. L. Pundir and Amrita, "Feature Selection Using Random Forest in Intrusion Detection," *Int. J. Adv. Eng. Technol.*, vol. 6, no. 3, pp. 1319–1324, 2013.

[24] A. A. Ramaki, M. Khosravi-Farmad, and A. G. Bafghi, "Real time alert correlation and prediction using Bayesian networks," *2015 12th Int. Iran. Soc. Cryptol. Conf. Inf. Secur. Cryptol.*, vol. 978, pp. 98–103, 2015.

[25] R. Kaur and M. Sachdeva, "International Journal of Advanced Research in An Empirical Analysis of Classification Approaches for Feature Selection in Intrusion Detection," vol. 6, no. 9, 2016.

[26] J. H. Assi and A. T. Sadiq, "NSL-KDD dataset Classification Using Five Classification Methods and Three Feature Selection Strategies," vol. 7, no. 1, pp. 15–28, 2017.

[27] Kamesh and N. Sakthi Priya, "Security enhancement of authenticated RFID generation," *Int. J. Appl. Eng. Res.*, vol. 9, no. 22, pp. 5968–5974, 2014.

[28] M. K. Gambo and A. Yasin, "Hybrid Approach for Intrusion Detection Model Using Combination of K-Means Clustering Algorithm and Random Forest Classification," *Ijes*, vol. 6, no. 1, pp. 93–97, 2017.

[29] Dubb Shruti and Sood Yamini, "Feature Selection Approach for Intrusion Detection System," *Int. J. Adv. Trends Comput. Sci. Eng.*, vol. 2, no. 5, pp. 47–53, 2013.

[30] N. Shahadat, I. Hossain, A. Rohman, and N. Matin, "Experimental Analysis of Data Mining Application for Intrusion Detection with Feature reduction," pp. 209–216, 2017.

[31] A. Jain and J. L. Rana, "Classifier Selection Models for Intrusion Detection System (Ids)," *Informatics Eng. an Int. J.*, vol. 4, no. 1, pp. 1–11, 2016.

[32] M. R. Parsaei, S. M. Rostami, and R. Javidan, "A Hybrid Data Mining Approach for Intrusion Detection on Imbalanced NSL-KDD Dataset," vol. 7, no. 6, pp. 20–25, 2016.