Hybrid feature selection technique for intrusion detection system

Muhammad Hilmi Kamarudin*, Carsten Maple and Tim Watson

Cyber Security Centre,
Warwick Management Manufacturing,
University of Warwick,
CV47AL, Coventry, UK
Email: hilmi_kamarudin@yahoo.com
Email: cm@warwick.ac.uk
Email: tm@warwick.ac.uk
*Corresponding author

Abstract: High dimensionality's problems have make feature selection as one of the most important criteria in determining the efficiency of intrusion detection systems. In this study we have selected a hybrid feature selection model that potentially combines the strengths of both the filter and the wrapper selection procedure. The potential hybrid solution is expected to effectively select the optimal set of features in detecting intrusion. The proposed hybrid model was carried out using correlation feature selection (CFS) together with three different search techniques known as best-first, greedy stepwise and genetic algorithm. The wrapper-based subset evaluation uses a random forest (RF) classifier to evaluate each of the features that were first selected by the filter method. The reduced feature selection on both KDD99 and DARPA 1999 dataset was tested using RF algorithm with ten-fold cross-validation in a supervised environment. The experimental result shows that the hybrid feature selections had produced satisfactory outcome.

Keywords: machine learning; filter-subset evaluation; wrapper-subset evaluation; genetic algorithm; random forest.

Reference to this paper should be made as follows: Kamarudin, M.H., Maple, C. and Watson, T. (2019) 'Hybrid feature selection technique for intrusion detection system', *Int. J. High Performance Computing and Networking*, Vol. 13, No. 2, pp.232–240.

Biographical notes: Muhammad Hilmi Kamarudin received his BSc in Computer Network from the Universiti Putra Malaysia, Selangor, Malaysia in 2007, and MSc in Computer Network from the Universiti Teknologi Mara, Selangor, Malaysia in 2010. He is currently pursuing his PhD in Computer Science with the University of Warwick, Coventry, UK. His research interests include network security, digital forensics, machine learning, and data mining.

Carsten Maple is a Professor of Cyber Systems Engineering. He is the Director of Research in Cyber Security working with organisations in key sectors such as manufacturing, healthcare, financial services and the broader public sector to address the challenges presented by today's global cyber environment. He has an international research reputation and extensive experience of institutional strategy development and interacting with external agencies. He has published over 200 peer reviewed papers and is co-author of the UK Security Breach Investigations Report 2010, supported by the Serious Organised Crime Agency and the Police Central e-crime Unit. He is also the co-author of *Cyberstalking in the UK*, a report supported by the Crown Prosecution Service and Network for Surviving Stalking. Additionally, he has advised executive and non-executive directors of public sector organisations and multi-billion pound private organisations.

Tim Watson is a Professor of Cyber Security Centre at the WMG. With more than 20 years of experience in the computing industry and in academia, he has been involved with a wide range of computer systems on several high-profile projects and has acted as a consultant for some of the largest telecoms, power and oil companies. He has designed, produced and delivered innovative courses on cyber security for a variety of public and private-sector organisations. His current research includes EU funded projects on combating cyber crime and research into the protection of infrastructure against cyber attack. He is the Vice President (Academia) of the Trustworthy Software Initiative, a UK Government sponsored project to make software better, and a key deliverable of the UK National Cyber Security Programme. He is also a regular media commentator on digital forensics and cyber security.

This paper is a revised and expanded version of a paper entitled 'Hybrid feature selection technique for intrusion detection system' presented at TRUSTCOM2016, Tianjin, China, 24 August 2016.

1 Introduction

The rapid growth of high-speed internet has led to an increase in data production at a staggering pace. Research by Koff and Gustafson in 2011 predicted that data assembly will be 44 times greater in year 2020 than it was in year 2009. Currently, the ability to analyse the increased data size is still not commensurate with its growth rate. When analysing voluminous data, it is obvious that more computational efforts are needed due to its sophistication and complexity.

The need for longer processing time would significantly affect the system performance as it slows down the attack detection speed. Thus, machine-learning tools was introduced to analyse the big (high dimensional) data. The key function of the tools is feature selection. The machine would recognise the most prominent feature for continues learning. It focuses on the learning algorithm to gather the most useful information for the future prediction requirement. Meanwhile, the system would still maintain its important features of primary data with better processing time.

The intrusion detection system (IDS) can be described as a device or an application that detects malicious activities or policy violations within the network. IDS have been widely used in recent years as one of the main network security components. The objective of this study is to find the best-fit approach that would significantly reduced the number of features. In addition, the approach would lead to high classification accuracy with less processing time.

To achieve this objective, we propose a hybrid feature selection model that leverages strengths from the legacy filter and wrapper selection procedure. The proposed hybrid solution is expected to effectively select the optimal set of features in detecting intrusion. The rest of this paper is organised as follows. Section 2 describes the background of feature selection, while the related work is explained in Section 3. The proposed method and results are presented in Sections 4 and 5. Section 6 concludes this finding and outlines future work.

2 Feature selection

Feature selection is a foundation of machine learning that has been explored for many years (Liu and Motoda, 1998). It is the process of discovering the most prominent feature for the learning algorithm. The most useful data is used for analysis to achieve better future projection. Therefore, the redundant or irrelevant features need to be extracted or removed to prevent the classifier from being biased towards more frequent recording. As the effectiveness of the algorithm selected is highly dependent on the feature selection, it is imperative to minimise the selecting features errors that could reduce the detection of abnormal behaviour. Choosing the feature selection algorithm often requires expert knowledge as it is not an easy task to determine a good set of features. Basically, there are two general methods namely filters and wrappers (Johnson and Shanmugam, 2011) that currently being used in many feature-selection processes.

The two categories of filters method include filter-based feature ranking (FBFR) and filter-based subset evaluation (FBSE). FBFR ranks the applicable features by assigning weights to individual features based on the score of every single feature to the target classes; no attention is spared to interaction between features. Feature ranking is faster than filter subset as it only computes the features once. In comparison, the FBSE computes to the power of two or 2n where (n = number of features). In view that the filter ranking features are uniquely selected, it did not take into account the relationship between features. As such, it was not able to handle redundant features effectively (Tang et al., 2014). Information gain (IG), mutual information and gain ratio are examples of FBFR.

FBSE was introduced simply to overcome the redundant features issues. It examines the whole subset in a multivariate way. It selects the relevant features and explores the degree of relationship between features. As such, selecting features in IDS, FBSE is more desirable than the FBFR (Nguyen et al., 2010). FBSE is heuristic-based and involves probabilities and statistical measures to search and evaluate the usefulness of all identified features. On the other hand, the wrapper-based subset evaluation (WBSE) uses a classifier to evaluate the worth of each feature subset. Usually, WBSE has better predictive accuracy compared to filters. This is because the selection approach is optimised when evaluating each feature subset with a particular classification algorithm. Conversely, most of the time wrappers are using classification algorithm to evaluate each set of features. This has made it excessively expensive to run. Moreover, when dealing with a large database that consists of many features (Hall, 1999) wrapper can become uncontrollable. Wrappers are also highly associated with the classifier's algorithm that makes it more difficult when shifting from one classifier to another. This is because the selection process needs total re-initiation.

Unlike filters, the selection criteria of features use distance measures and correlation function (Cleetus, 2014). It does not require re-execution for different learning classifiers. This has made its execution much faster than the wrappers. Filters are suitable in large database environment that contains many features. Researchers have often used the filter, as an alternative to the wrapper since the latter is expensive and time-consuming to run.

2.1 Correlation-based feature selection

In this study, we use correlation-based feature selection (CFS) that derived from Pearson correlation coefficient. The CFS is a simple filter algorithm that evaluates subsets of feature according to heuristic evaluation function. The study was based on the hypothesis "A good feature subset is one that contains features highly correlated with the class, yet uncorrelated with each other" (Hall, 1999). The algorithm will remove irrelevant features that have low correlations with the class. It would also screen out all redundant features as they might highly correlate with other features. The redundant feature occurs when one or more features are highly correlated with each other. The following equation from (Hall, 1999) shows how the M merit is used to select subset S containing k number of features. Both redundant and irrelevant features are determined from $\overline{rcf}$ where it is the μ, mean of correlation for each feature and its class while $\overline{rff}$ is the μ, mean of correlation between features.

$$MS = \frac{krcf}{\sqrt{k + k(k-1)rff}} \tag{1}$$

The equation concludes that the probability of features to be selected will depend on the correlation between feature and the class as well as correlation among the features. The process would be continued explore the search space with heuristic search algorithm. The subsets with the highest merit obtained during the search would be selected.

2.2 Random forest classifier

The random forest (RF) algorithm can be classified as an ensemble classification and regression tree (CART). This algorithm is widely used in data mining techniques for prediction, pattern recognition and probability estimation (Zhang et al., 2008; Attal et al., 2015; Khoshgoftaar et al., 2007). It consists of many decision tree classifiers. Each decision tree is constructed from a different sample of the original dataset. The outputs are chosen based on a vote from each tree that indicated the tree's decision of the class object. The most votes for the object are from the best individual trees.

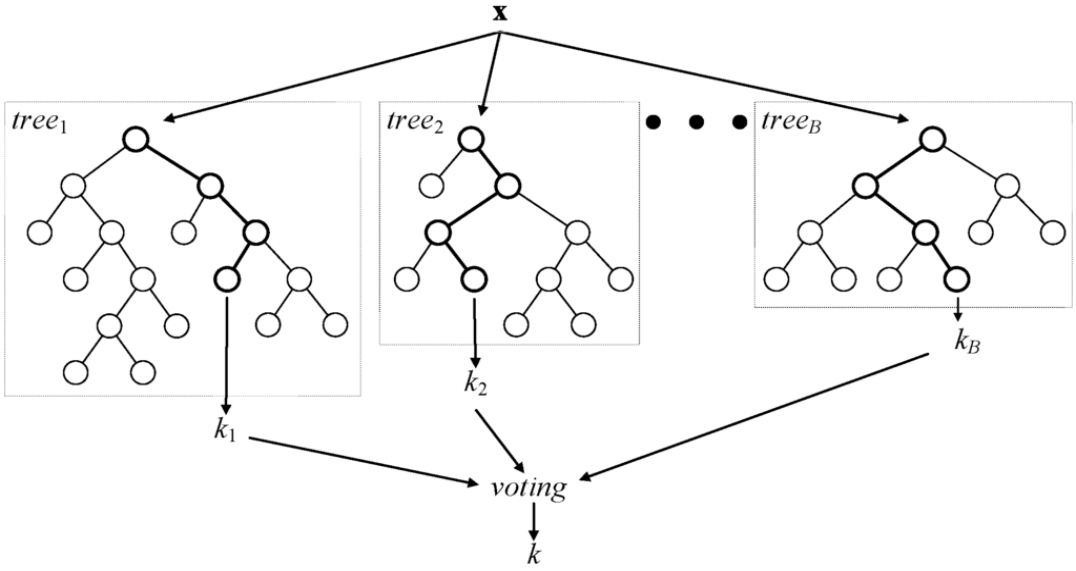
As the RF generates many classification trees, each tree is constructed by a different bootstrap sample from the original data using a tree classification algorithm. Each tree gives a vote that indicates the tree's decision about the class of the object. The forest chooses the class with the most votes for the object. Out-of-bag (OOB) error is used as validation during the tree growth. It is described as the average of the classification error that is connected to each tree *Tb* using the *OOBb* sample. After the forest is formed, a new sample *xi* needs to be classified as per the following equation:

$$\hat{C}_{rf}^{B}(xi) = majority\ vote\{Cb(xi)\}_{1}^{B} \tag{2}$$

where $\hat{C}b(xi)$ is the class that assigned by the tree *Tb*.

RF as a classification technique has been extensively used due to its robustness in handling high dimensional data (Hastie et al., 2009) with small number of available learning samples (Htun and Khaing, 2013). It also runs on nominal data and resolves over-fitting drawback issues effectively.

Figure 1 A general architecture of a RF



2.3 Heuristic search

The search algorithm performance is important in selecting the best-fit features. One of the popular search techniques used is called the greedy algorithm. The features are added or removed from the subset until the algorithm has considered all possible selections. When the algorithm is added to the feature subset, it is recognised as 'forward selection'. On the other hand, when the feature subset is deleted, it is called as 'backward elimination' (Guyon, 2003). This approach was not a popular choice as could suffering from the 'nesting effect'. Furthermore, when features are removed from top-down, the algorithm is unable to place them back onto the selected subset. As such, the final subset may not possess the best features. The final subset would only be considered when it has shown better performance over than the previous subset (Wald et al., 2013a). This approach indicates strong assumption that the features are completely independent when using evaluation metric on an individual feature (Diao and Shen, 2012).

The best-first search algorithm is an artificial intelligence approach that allows a backtracking search. Similar to greedy algorithm, the best-first search technique is also used to explore the most promising path by moving through the search space and making changes to the current feature subset. The most recent subset is checked at each step. If it is found to be better than the previous subset, then the recent will replace the previous subset (Wald et al., 2013a). Unlike the greedy algorithm, the best-first algorithm can go back to the previous subset and continue from there if the search begins to look less promising. The best-first entire space exploration characteristic has made it common to stopping criterion. It normally limits the number of expended subsets that return without any improvement.

Genetic algorithm (GA) is an adaptive search technique that uses biological evolution as natural selection principle in solving a problem (Hassan, 2013). It is based on the evolution theory proposed by Darwin. According to Darwin, the survival of the fittest is the concept to converge optimal solution and make effective population. The algorithm applies iterative process that evolves when applying selection, crossover and mutation operators to the members of the current generation (Bagyamani et al., 2013). The selection process will identify the fittest subset of the current population to serve as a parent for the next generation. Crossover operators are a primary exploration mechanism for GA, combining different features from a pair of subsets to form a new subset. The operation will takes two feature subsets (parents) and reconstruct the two new subsets (child's). The mutation operator's main function is to restore the diversity that may be lost during the repeated selection and crossover application. It modifies certain value in a subset randomly. During the evolution search, the fittest of the subset is estimated using the fitness function. Using the earlier three operator's processes, a better fitness feature subsets would have greater chance of being selected and forming a new subset. The genetic search gives a global optimum solution and is more robust compared to the greedy approach and best-first. However, it does come with some computational effort (Vafaie and Imam, 1994).

3 Related work

Feature selection process has attracted interest of many researchers due to its potentiality in reducing high dimensional data. Feature ranking algorithm was introduced merely to select the top six features based on rank (Sung and Mukkamala, 2004). The authors used three ranking algorithms of support vector machines (SVMs), multivariate adaptive regression splines (MARS), and linear genetic programming (LGP). The algorithm would select the best feature and make performance comparison between each algorithm. The detection would be programmed to detect Probe and DoS attack. The LGP could achieve higher accuracy rate in detecting both types of attacks compare to other algorithms. This approach is however effective for specific type of attacks only. It is not fit for others such as R2L and U2R attacks.

The speedy computation ability of the filter ranking has made it suitable for very large datasets. For instance, Wald et al. (2013a) use filter ranking to reduce 480 to 40 features. They compare three different approaches of feature selection (filter-rank, filter-subset evaluation and wrapper-subset evaluation) to find the best method to select the relevant features. Three different feature selections with six different classifiers, five-nearest neighbours (5-NN), logistic regression (LR), multi-layer perceptron (MLP), naive Bayes (NB), RF with 100 trees (RF100), and SVMs were used to achieve the best results. As recommended by the authors, the filter ranking process executed in high dimension data performs better when using SVM classifier techniques compare to the other two. Nevertheless, the authors also proved that the filter ranking method is more competent than the filter subset and the wrapper subset. However, there was no explanation on the methods implemented in choosing the top 40 features from the ranking table. This uncertainty might affect the final optimal set of features since it may contain irrelevant features.

Another filter ranking was implemented by Ambusaidi et al. (2014), who proposed hybrid feature selection by combining both mutual information (filter ranking) and wrapper that using least square-SVM as classification algorithm in removing irrelevant and redundant features. Mutual information provides a good measurement to find relevant feature by quantify the amount of information to the output class. The proposed criterion function G complements mutual information in removing redundant features. Although the results shows significant reduction of features, the false positive and the detection rate still can be improved.

In relation to the calculation of features selection relevancy, El-Khatib (2010) has proposed information gain ratio (IGR) to replace the IG calculation method. This is because the IG is normally biased towards features that have high distinct value. The selected features are ranked based on score derived from the IGR calculation. The K-means

classifier is then used to determine the best-fit feature-set based on performance results accuracy. The selection process would end when the current subset performance drops below the previous subset accuracy. The selected features are tested with three types of artificial neural network (ANN) architecture namely perceptron, multilayer backpropagation perceptron (MBP) and hybrid multilayer perceptron (HMP). Although HMP has lower false positive rate and takes longest time in learning model, its classifier has outperformed both the perceptron and the MBP. Nevertheless, there were no significant detection rate differences between the proposed HMP and MBP.

It proved less accurate when classifier was used singly to evaluate performance accuracy compared to using ensemble technique (combining more than one classifier) (Mukkamala et al., 2004). Zainal et al. (2008) proposed neural fuzzy inference system (ANFIS) and LGP algorithms in detecting four main types of attacks of Probe, DoS, R2L and U2R. This ensemble technique was implemented with a reduced set of features (between six to eight only) for each type of attacks. This technique has achieved more than 99% detection rate for R2L and U2R attack types and an average of 99.15% accuracy for all attack types.

In selecting the best-fit classifier (Zaman and Karray, 2009), the authors compared thier approach with two different classifiers namely neural network (NN) and SVM. They proposed a novel method called enhanced support vector decision function (ESVDF) to select features based on rank and use backward elimination ranking (BER) and forward selection ranking (FSR) to calculate the correlation between features. The comparison of both algorithms reveals that the NN had better performance than the SVM, with 99.55% accuracy. The method had improved in reducing the number of features and the time taken to build a model, by a negligible margin of 0.08% and 0.11% for NN and SVM, respectively. Nevertheless, the accuracy rate is still lower than when using full features.

Numeral feature selection approaches had been introduced in the past. Nevertheless, achieving low false detection and high attack recognition capabilities is still a major challenge. There are three main differences in our approaches compared to existing hybrid selection (Ambusaidi et al., 2014; Singh and Tiwari, 2015). The first approach uses correlation-based selection to determine the worthiness of each feature by removes redundant features that exist inside filter rank. The second approach analyses and determines the best-fit search strategy. Here, the top three search algorithms of 'best-first', 'greedy' and 'genetic search' are concisely evaluated. The third approach uses RF as a classification algorithm in the evaluation process. This is due to its effectiveness for high dimensional data execution, complemented with capability of solving over-fitting issues (Hastie et al., 2009). The concept of our proposed hybrid feature selection technique was based on leveraging the strengths of both the filter and the wrapper techniques.

4 The proposed hybrid feature selection technique

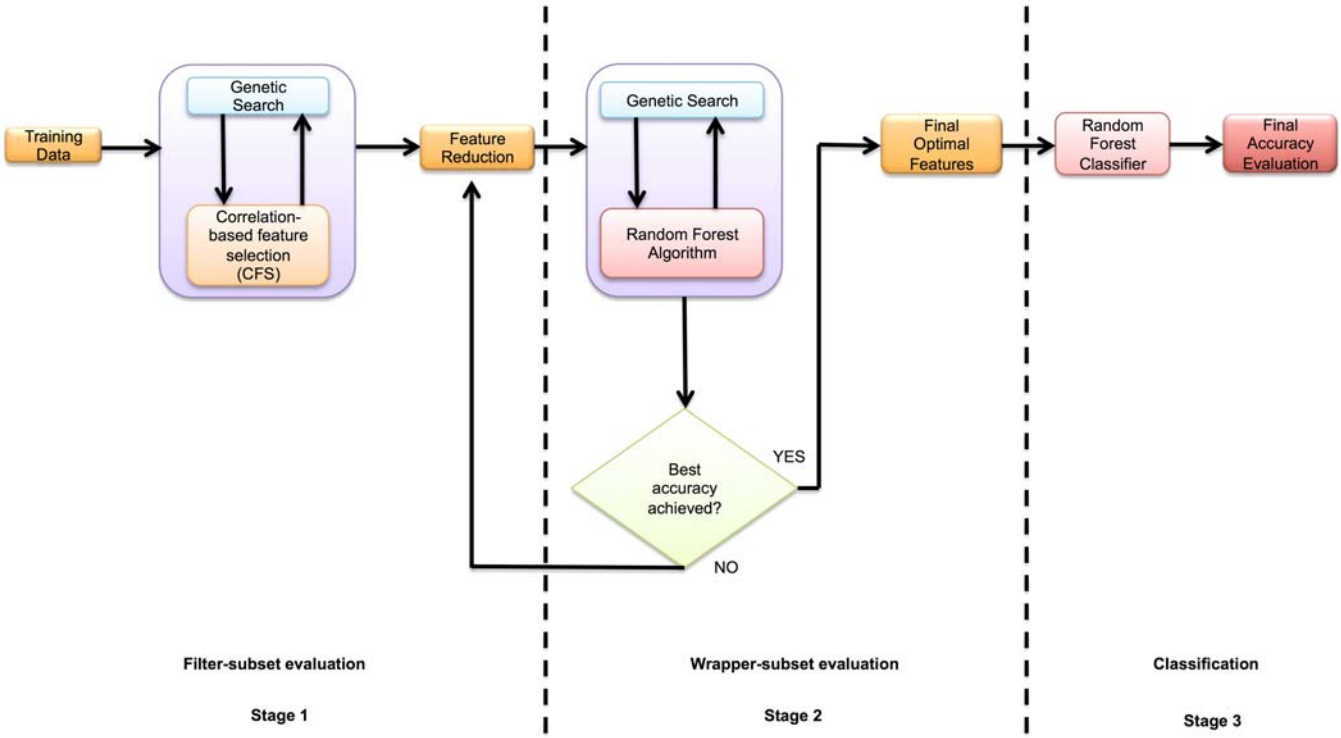
Figure 2 shows the process flows for building hybrid feature selection. The process is divided into three stages as follows:

In stage 1, the process starts with the filter-subset evaluation. It processes the original features M to generate new set L of reduced features, where $L \subseteq M$. We proposed the CFS approach due to its robustness in removing redundant and irrelevant features. In this research, CFS is use to prevails over problems in feature ranking (IG, gain ratio) where redundant features exist (Wald et al., 2013a) and reduced features are usually defined without performing further examination. We are not considering exhaustive search because it is proven as not suitable to use in large dataset (Guyon 2003) due to its high complexity. As such, we had applied heuristic search techniques and chose a GA as the search function. This technique had demonstrated better performance during our experiments between the best-first and the greedy search methods, refer to Tables 3 and 4. Furthermore, in this stage, it is crucial to help truncate the computational effort for wrapper approach, which only deals with reduce feature set compared to the original set of features.

In stage 2, the reduced feature set L gathered from the FBSE had been continued with WBSE to produce the final optimal features K, where $K \subseteq L \subseteq M$. The hybridisation of both the filter and wrapper was proposed due to the filter approach alone would not able to find the best available subset as it is less dependent on the classifier (Peng et al., 2010). On the other hand, the wrapper approach is believed to be more effective and to produce better accuracy. However, it is computationally expensive when dealing with large dataset. In view of the above limitations, we had leveraged the strength of both methods to form a better-synergised approach. In WBSE, we use the RF classifier to evaluate the selected features using the genetic search and produced the final K feature subset. The search will continue to train new model for each subset and will stop when the final optimum subset has been found.

Stage 3 is called the classification stage. In this stage, the final optimum subset K, produced by WBSE, was tested by the RF classifier with ten-fold cross-validation. Eight performance metrics were used in this experiment and the output indicates that the hybrid feature selection approach yielded better results compared with other approaches.

Figure 2 A proposed hybrid feature selection design (see online version for colours)



5 Experiments and results

5.1 Experimental setup

These experiments have been obtained using WEKA data mining tool (Anon, n.d.) version 3.7.1 on Pentium core i7 in Windows environment.

5.2 Network and host-based dataset

We applied KDD99 dataset (network-based) and Defense Advanced Research Projects Agency (DARPA) 1999 dataset (host-based) to test the proposed methods in two different environments. Initially, we were quite sceptical on the DARPA dataset (McHugh, 2000) perfections especially on its maturity over 15 years. Later, we found out that it was actually the most comprehensive and extensively used dataset. In fact, the DARPA dataset has been recognised as the standard benchmark by many researchers in this field (Xiang and Zhou, 2006; Yassin et al., 2014).

The KDD99 intrusion detection dataset is based on the DARPA 1998 initiative that has created a benchmark in evaluating different methodologies. DARPA 1999 was the improved version of DARPA 1998, as it contains more new attacks than the previous dataset. In DARPA 1999 dataset, we had selected a host with the IP address 172.016.112.050 because it had the most number of attacks among hosts within the dataset (Shamsuddin and Woodward, 2007).

Table 1 shows the description of both dataset used in this experiment.

KDD99 training dataset consists of 494,014. Out of which only 19.6% or 97,271 of the data were recognised as normal data. The attack was divided into four categories namely, Probe (4,107), DoS (391,458), U2R (59) and R2L (1,119) (Jalil et al., 2010). In DARPA 1999 dataset, 87.6% or 465,409 were representing normal data while the remaining 12.4% or 65,821 were attack data.

Table 1 KDD99 training dataset

Class name	*Instance*	*Percentage %*
Normal	97,271	19.6
Attack (DoS, Probe, U2R and R2L)	396,743	80.4

Table 2 DARPA 1999 (host-based) 172.016.112.050

Class name	*Instance*	*Percentage %*
Normal	465,409	87.6
Attack (DoS, Probe, U2R and R2L)	65,821	12.4

5.3 Results and analysis

During the experiment, we observed that both filter and hybrid methods had selected the same attributes (f6, f23, f31, f37) in KDD99 and (f28, f31, f32) in DARPA1999. We found that this is due to the obvious attack pattern demonstrating by those features. For instance, feature 23 (count: number of connections going into the same host in the past two seconds) has indicated a significant pattern of Probe and DoS attacks. The nature of these attacks was targeting a specific host by sending huge traffic volume to the same host to flood the whole network and used-up all

available resources. Unlike the U2R and R2L pattern, the attack behaviour usually takes longer time to access the system to guess the password or by brute force attack. This will make feature 1 (duration: number of seconds of the connection) more important in detecting these types of attacks. In this research process, we found that the greedy and best-first approaches select the same features. This was due to their nature of sharing similar space search techniques. On the other hand, GA those pick-up different features should provide global optimum solution.

Tables 3 and 4 show comparison of performance metrics among the three different search methods (best-first, greedy-stepwise and genetic search) on FBSE and hybrid approach using KDD99 and DARPA 1999 datasets. Out of the three search methods, genetic search has scored the highest accuracy rate of 0.42% more than the other two search methods. We then hybridise the reduced features from filter with wrapper RF classifier. Although filter method could also show some performance improvement, but when combined with wrapper approach, it had recorded significant features reduction from 21 to 12 and improved the time building model by 14%. Moving on, we had tested the hybrid method on DARPA 1999 host-based environment. The result shown was significant reduction of features from 33 to 5 with higher accuracy and detection rate by 1.22 % and 0.8%. The results clearly demonstrated the accuracy detection is either the same or better with considerable reduction of features set. Thus, it is an essential step to utilise feature selection techniques in building IDS.

5.4 Performance metrics

In this study, we have considered several performance metrics (number of features, time, true positive (TP), true negative (TN), false positive (FP), false negative (FN), detection rate and accuracy) as a benchmark to evaluate the proposed method (Wald et al., 2013b).

a Number of features: Feature used during experiment.

b Time: The time measured in seconds taken by classifier to build the model on dataset.

c TP: To estimate the amount of attack data detected is actually attack data.

d TN: To estimate the amount of normal data detected is actually normal data.

e FP: To quantify the amount of normal data detected as attack data.

f FN: To quantify the amount of attack data detected as normal data.

g Detection rate: Is the proportion of detected attacks among all attack data.

h Accuracy: Measured as a percentage, where instances are correctly predicted.

$$Detection\ rate\ (DR) = \frac{(TP)}{(TP)+(FN)} \tag{3}$$

$$Accuracy\ (ACC) = \frac{(TP)+(TN)}{(TP)+(TN)+(FP)+(FN)} \tag{4}$$

Table 3 Performance metrics of different feature selection approaches for KDD99 (network-based) dataset

KDD99 dataset (network-based)								
	Number of features	*Time*	*True positive*	*True negative*	*False positive*	*False negative*	*Detection rate*	*Accuracy*
Full feature	41	122.71 s	99.98%	99.97%	0.03%	0.01%	99.99%	99.98%
Filter (best first)	6	11.94 s	99.57%	99.53%	0.47%	0.43%	99.57%	99.56%
Filter (greedy stepwise)	6	11.94 s	99.57%	99.53%	0.47%	0.43%	99.57%	99.56%
Filter (genetic search)	21	16.27 s	99.99%	99.97%	0.03%	0.01%	99.99%	99.98%
Proposed method (filter + wrapper)	12	13.98 s	99.99%	99.97%	0.03%	0.01%	99.99%	99.98%

Table 4 Performance metrics of different feature selection approaches for DARPA 1999 (host-based) dataset

DARPA 1999 (host-based)								
	Number of features	*Time*	*True positive*	*True negative*	*False positive*	*False negative*	*Detection rate*	*Accuracy*
Full feature	33	24.24 s	98.71%	99.81%	0.19%	1.29%	98.71%	99.68%
Filter (best first)	6	21.61 s	98.71%	99.95%	0.05%	2.16%	97.85%	99.69%
Filter (greedy stepwise)	6	21.61 s	98.71%	99.95%	0.05%	2.16%	97.85%	99.69%
Filter (genetic search)	8	22.43 s	99.24%	99.94%	0.06%	0.76%	99.24%	99.85%
Proposed method (filter + wrapper)	5	17.12 s	99.51%	99.99%	0.01%	0.49%	99.51%	99.93%

5.5 Respective study

Table 5 shows comparison between our hybrid feature selection with other feature selection methods that use similar KDD99 dataset, we noted that it produced better accuracy rate. Thus, our hybrid feature selection method has proven a better option in detecting network and host traffic intrusion.

Table 5 Comparison results between our approach and others feature selection approach on KDD99 dataset

Methods	*Features*	*False positive*	*Detection rate*	*Accuracy*
Enhanced support vector decision function (Zaman and Karray, 2009)	9	0.03%	-	99.58%
Generic feature selection (Nguyen et al., 2010)	18	-	-	99.60%
Hybrid classification algorithm (Singh and Tiwari, 2015)	20	-	-	99.70%
Least square SVM (Ambusaidi et al., 2014)	6	0.07%	99.93%	99.90%
Proposed method	12	0.03%	99.99%	99.98%

6 Conclusions and future works

The objective of the research was to capitalise on the strengths of both the filter and the wrapper approaches as the pre-processing phase in detecting intrusion. We tested the hybrid feature-selection model by combining the strengths of the two FBSE and WBSE approaches. Initially, the FBSE was to concentrate on reducing the WBSE computational effort by filtering the insignificant and redundant features. Then it continues searching for the optimal subset that was first picked up by the FBSE to improve the classification performance. The final features subset generated from the hybrid process was tested using RF classifier and ten-fold cross-validation. The FBSE was also used to resolve the filter ranking issue of existing redundant features. The proposed hybrid feature selection was evaluated with two types of datasets (network-based and host-based) mainly to allow different integration testing environment. Individually, FBSE approach is capable enough to demonstrate some performance improvement. However, the newly generated hybrid combination model of FBSE and WBSE approaches yielded better performance improvements in terms of detection time, accuracy, and detection rate. It also recorded a low false alarm rate. Moving forward, we will proceed with the research on the ability of this newly discovered hybrid model in detecting novel attacks.

References

Ambusaidi, M.a. et al. (2014) 'A novel feature selection approach for intrusion detection data classification', *2014 IEEE 13th International Conference on Trust, Security and Privacy in Computing and Communications*, pp.82–89.

Anon (n.d.) *Machine Learning 'WEKA'* [online] http://www.cs.waikato.ac.nz/ml/weka (accessed 1 October 2016).

Attal, F. et al. (2015) 'Powered two-wheeler riding pattern recognition using a machine-learning framework', *IEEE Transactions on Intelligent Transportation Systems*, Vol. 16, No. 1, pp.475–487.

Bagyamani, J., Thangavel, K. and Rathipriya, R. (2013) 'Biclustering of gene expression data based on hybrid genetic algorithm', *International Journal of Data Mining, Modelling and Management*, Vol. 5, No. 4, p.333.

Cleetus, N. (2014) 'Genetic algorithm with different feature selection method for intrusion detection', *First International Conference on Computational Systems and Communications (ICCSC)*, December.

Diao, R. and Shen, Q. (2012) 'Feature selection with harmony search', *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, Vol. 42, No. 6, pp.1509–1523.

El-Khatib, K. (2010) 'Impact of feature reduction on the efficiency of wireless intrusion detection systems', *IEEE Transactions on Parallel and Distributed Systems*, Vol. 21, No. 8, pp.1143–1149.

Guyon, I. (2003) 'An introduction to variable and feature selection', *Journal of Machine Learning Research*, Vol. 3, No. 3, pp.1157–1182.

Hall, M.a. (1999) *Correlation-based Feature Selection for Machine Learning*, University of Waikato [online] http://www.cs.waikato.ac.nz/~mhall/thesis.pdf.

Hassan, M.M. (2013) 'Network intrusion detection system using genetic algorithm and fuzzy logic', *International Journal of Distributed and Parallel Systems (IJDPS)*, Vol. 4, No. 2, pp.1435–1445.

Hastie, T., Tibshirani, R. and Friedman, J. (2009) 'The elements of statistical learning: data mining, inference, and prediction', *The Mathematical Intelligencer*, Vol. 27, No. 2, pp.83–85.

Htun, P.T. and Khaing, K.T. (2013) 'Detection model for denial-of-service attacks using random forest and k-nearest neighbors', *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)*, Vol. 2, No. 5, pp.1855–1860.

Jalil, K.a., Kamarudin, M.H. and Masrek, M.N. (2010) 'Comparison of machine learning algorithms performance in detecting network intrusion', *2010 International Conference on Networking and Information Technology ICNIT*, pp.221–226.

Johnson, S. and Shanmugam, V. (2011) 'Effective feature set construction for SVM-based hot method prediction and optimisation', *International Journal of Computational Science and Engineering*, Vol. 6, No. 3, p.192.

Khoshgoftaar, T.M. et al. (2007) 'Estimating class probabilities in random forest', *Proceedings – 6th International Conference on Machine Learning and Applications, ICMLA 2007*, pp.348–353.

Koff, W. and Gustafson, P. (2011) 'CSC leading edge forum data revolution', *CSC Leading Edge Forum*, p.68.

Liu, H. and Motoda, H. (1998) *Feature Selection for Knowledge Discovery and Data Mining*, Kluwer Academic Print Publisher.

McHugh, J. (2000) 'Testingintrusion detection systems: a critique of the 1998 and 1999 DARPA intrusion detection system evaluations as performed by Lincoln Laboratory', *ACM Transactions on Information and System Security*, Vol. 3, No. 4, pp.262–294.

Mukkamala, S., Sung, A.H. and Abraham, A. (2004) 'Modeling intrusion detection systems using linear genetic programming approach', *Innovations in Applied Artificial*, Vol. 1, pp.633–642.

Nguyen, H.T., Petrović, S. and Franke, K. (2010) 'A comparison of feature-selection methods for intrusion detection', *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), LNCS*, Vol. 6258, pp.242–255.

Peng, Y., Wu, Z. and Jiang, J. (2010) 'A novel feature selection approach for biomedical data classification', *Journal of Biomedical Informatics*, Vol. 43, No. 1, pp.15–23.

Shamsuddin, S.B. and Woodward, M.E. (2007) 'Modeling protocol based packet header anomaly detector for network and host intrusion detection systems', *Proceedings of the 6th International Conference on Cryptology and Network Security*, pp.209–227.

Singh, P. and Tiwari, A. (2015) 'An efficient approach for intrusion detection in reduced features of KDD99 using ID3 and classification with KNNGA', *Proceedings – 2015 2nd IEEE International Conference on Advances in Computing and Communication Engineering, ICACCE 2015*, pp.445–452.

Sung, A.H. and Mukkamala, S. (2004) 'The feature selection and intrusion detection problems', *Advances in Computer Science – Asian 2004, Proceedings*, Vol. 3321, pp.468–482.

Tang, J., Alelyani, S. and Liu, H. (2014) 'Feature selection for classification: a review', *Data Classification: Algorithms and Applications*, pp.37–64.

Vafaie, H. and Imam, I.F. (1994) 'Feature selection methods: genetic algorithms vs. greedy-like search', *Proceedings of the International Conference on Fuzzy and Intelligent Control Systems*, March, Vol. 1.

Wald, R., Khoshgoftaar, T. and Napolitano, A. (2013a) 'Filter- and wrapper-based feature selection for predicting user interaction with Twitter bots', *Proceedings of the 2013 IEEE 14th International Conference on Information Reuse and Integration, IEEE IRI 2013*, pp.416–423.

Wald, R., Khoshgoftaar, T. and Napolitano, A. (2013b) 'The importance of performance metrics within wrapper feature selection', in *2013 IEEE 14th International Conference on Information Reuse & Integration (IRI)*, IEEE, pp.105–111 [online] http://ieeexplore.ieee.org/document/6642460/ (accessed 1 October 2016).

Xiang, Y. and Zhou, W. (2006) 'Protecting information infrastructure from DDoS attacks by MADF', *International Journal of High Performance Computing and Networking*, Vol. 4, Nos. 5/6, p.357.

Yassin, W. et al. (2014) 'Packet header anomaly detection using statistical analysis', *International Joint Conference SOCO'14-CISIS'14-ICEUTE'14 SE-47*, No. 299, pp.473–482.

Zainal, A., Maarof, M.a. and Shamsuddin, S.M. (2008) 'Data reduction and ensemble classifiers in intrusion detection', *2008 Second Asia International Conference on Modelling and Simulation (AMS)*, pp.591–596.

Zaman, S. and Karray, F. (2009) 'Features selection for intrusion detection systems based on support vector machines', *2009 6th IEEE Consumer Communications and Networking Conference*, pp.1–8.

Zhang, J., Zulkernine, M. and Haque, A. (2008) 'Random-forests-based network intrusion', *Man and Cybernetics*, Vol. 38, No. 5, pp.649–659.