A Framework for Efficient Network Anomaly Intrusion Detection with Features Selection

Hebatallah Mostafa Anwer
Department of Computer Science
College of Computing and Information Technology
Arab Academy for Science, Technology & Maritime Transport
Alex, Egypt
hebatallahma@student.aast.edu

Mohamed Farouk
Department of Computer Science
College of Computing and Information Technology
Arab Academy for Science, Technology & Maritime Transport
Alex, Egypt
mfarouk316@aast.edu

Ayman Abdel-Hamid
Department of Computer Science
College of Computing and Information Technology
Arab Academy for Science, Technology & Maritime Transport
Alex, Egypt
hamid@aast.edu

***Abstract*— An intrusion Detection System (IDS) provides alarts against intrusion attacks where a traditional firewall fails. Machine learning algorithms aim to detect anomalies using supervised and unsupervised approaches. Features selection techniques identify important features and remove irrelevant and redundant attributes to reduce the dimensionality of feature space. This paper presents a features selection framework for efficient network anomaly detection using different machine learning classifiers. The framework applies different strategies by using filter and wrapper features selection methodologies. The aim of this framework is to select the minimum number of features that achieve the highest accuracy. UNSW-NB15 dataset is used in the experimental results to evaluate the proposed framework. The results show that by using 18 features from one of the filter ranking methods and applying J48 as a classifier, an accuracy of 88% is achieved.**

Keywords*— *Intrusion detection system, Machine learning techniques, Features selection methods.

I. INTRODUCTION

An Intrusion Detection System is designed to detect an intrusion while it is in progress, or after it has occurred. The major functions performed by IDS are monitoring users and systems activity, auditing system configurations, recognizing known attacks, identifying abnormal activities, managing audit data, highlighting normal activities, correcting system configurations and storing information about intruders [1].

There are two types of intruders. The external intruders are unauthorized users of the machines they attack. Internal intruders have permission to access the system with some restrictions. It is essential to build effective intrusion detection systems for protecting information systems against such attacks [2].

IDSs are classified based on two distinct aspects which are network-based IDSs versus host-based IDSs. Network-based IDS uses a set of sensors to capture the network packets to analyze them and evaluate the information found in the network communications. Host-based IDS uses system logs and audit trails to evaluate the information found on a single or multiple host systems [3].

Intrusion detection systems can be classified as misuse IDSs versus anomaly-based IDSs. Misuse IDS is a signature-based IDS which can detect known attacks in an efficient way based on hard coded signatures stored in the signature list. The misuse techniques have the advantage of low false positive rate. However, they suffer from high false negative rate due to the sensitivity to any simple variation in the stored signatures. In such case, the variations can be considered as an attack. Misuse IDS fails in detecting unknown and zero-day attacks where they are unavailable in the stored signatures [3].

Anomaly-based approaches use machine learning techniques to establish a normal profile usage. An anomalous request will be considered as an attack if it violates such normal profile. Supervised and unsupervised techniques are used to establish that profile resulting in low false negative rate. Anomaly-based approaches succeed in detecting unknown and zero-day attacks which is an advantage over the signature-based approaches. However, these techniques suffer from high false positive rate in such case of dealing with high dimensional datasets in the training process [3].

This paper presents a features selection framework that applies different features selection strategies. The framework uses machine learning algorithms where it is applicable for any dataset. The aim of this framework is to get the minimum number of features that achieve the highest accuracy under best performance. It is applied in a case study that uses five main strategies by using filter and wrapper methods with six single attribute evaluators and two machine learning classifiers based on UNSW-NB15 dataset for network intrusion detection.

The rest of this paper is organized as follows. Section II introduces the background overview and related work. Section III presents the framework proposed strategies. Section IV presents the experimental results. Finally, the paper is concluded in section V along with ideas for future work.

II. BACKGROUND AND RELATED WORK

Machine learning is a set of computational methods using example data or experience to improve performance, to make accurate predictions in the future and gain knowledge from data. There are steps to develop machine learning applications. These steps are collecting data, preparing the input data, analyzing the input data, training the algorithm, testing the algorithm and finally using it. Examples of such machine learning techniques are Decision Tree and Naïve Bayes [4].

978-1-5386-4366-2/18/$31.00 ©2018 IEEE

Features selection is an important pre-processing stage on datasets to be used in machine learning. Features selection reduces the dimensionality of data and enhances the performance of the classification process. Some of the features selection methods are Wrapper and Filter.

Wrapper method has been applied in many research areas for features selection by evaluating a subset of features obtained from training and testing a classification model. Possible features subsets are grouped for evaluation by applying a search procedure. This process needs to be repeatedly done each time a different classifier is used which can be considered as a drawback. A heuristic search can be used for this purpose to find the optimal subset where the space of features subsets grows exponentially [5].

Filter methods use a statistical model to score and rank the features depending on the intrinsic properties of the data. The features are sorted from the highest to the lowest ranked. These methods have the advantage of being fast as they are independent from the classifier and scalable with high-dimensional data. However, they ignore the correlation between features and the interaction with the classifier. The filter needs to be applied once on the dataset where different classifiers can be evaluated later. A threshold point can be found for cutting down the number of features which should have the lowest ranks [5]. There are different ranking evaluators for features selection. Some of the evaluators, as referenced in [6] are Info Gain (IG), Gain Ratio (GR), Symmetrical Uncertainty (SU), Relief F (RF), One R (OR) and Chi Squared (CS)

Section A discusses some of the references that applied machine learning algorithms in IDS. Other researchers in section B, applied some features selection methods before applying machine learning.

A. Machine learning techniques in IDS

In [7], four machine learning techniques are applied individually on the UNSW-NB15 and ISOT datasets to examine the performance and accuracy in the cloud security. These techniques are Decision Tree (J48), Support Vector Machine (SVM), Naïve Bayes (NB) and Logistic Regression (LR). Different datasets are used to test the robustness of the different classifiers. The conclusion is that a prevalent condition exists in cloud scenarios due to network function virtualization and service function chaining. A single dataset can't include all types of attacks. A supervised machine learning model that performs well with a particular dataset may not achieve a satisfactory performance with another. The paper suggested more research in the cloud security using supervised and unsupervised techniques.

In [8], the paper proposed a misuse detection model and an anomaly detection model as a hybrid intrusion detection system. At the first stage, the misuse detection model is achieved based on a Binary Tree of classifiers. Classifiers for known attacks are used to construct the binary tree. Each level consists of a certain classifier. For intrusion detection, the network information is checked level by level. If the classification fails in level (n), then it is passed on to the next classifier in level (n+1) until none are found. The misuse detection model will detect only known attacks. At the second stage, the anomaly detection model is achieved based on SVM classifier. The SVM is used as classifier for detecting patterns that deviate from normal behavior. This hybrid system aims to reduce the False Negative and False Positive rates. KDD Cup '99, NSL-KDD and UNSW-NB15 datasets are used to evaluate the proposed hybrid intrusion detection.

In [9], Multilayer Perceptron and Decision Tree are used to compare the intrusion detection system that can only access the packet headers of network traffic and not the attached data. The goal is to keep up with the growing bandwidth of networks and maintain the privacy of the users with one that can also access the payload data of network packets. The two different intrusion detection systems are compared for future adaptive cyber security and to analyze their performance as network traffic classification algorithms.

B. Features selection methods in IDS

In [10], the paper presents a victim-end DoS detection method based on Artificial Neural Networks (ANN) using a Feed forward back propagation learning algorithm. To detect DoS attack with minimum resources usage, unsupervised correlation-based features selection method is applied. The system consists of three steps which are, collection of the incoming network traffic, features selection for DoS detection and classification into DoS traffic or normal traffic. Two datasets UNSW-NB15 and NSL-KDD are used to evaluate the performance of this method. The results are satisfactory when compared to other DoS detection methods, where the features selection reduced the dimensionality of the data providing an improvement in the time of training and detection.

In [11], to decrease the False Alarm Rate (FAR) and improve the accuracy of detection, a features selection hybrid method based on the central points (CP) of attribute values followed by an Association Rule Mining (ARM) algorithm are used. CP is applied first on the dataset to divide it into equal partitions and hence reduce the processing time. The output is forwarded to the ARM to extract the highly ranked features. For network intrusion detection, Expectation-Maximization clustering, Logistic Regression and Naïve Bayes techniques are used in the decision engine for comparison. This algorithm used the UNSW-NB15 and the NSLKDD datasets.

In [12], a detection approach in cyber systems was presented based on a multimodal artificial neural network (MANN) using the collected network traffic data from completely observable cyber systems for training and testing. The genetic features selection algorithm is used to reduce the computational overhead. It is effective in solving optimization problems and handle multiple solution search spaces. Convolutional neural network (CNN) and k-means clustering algorithm are used as a supervised machine learning algorithm for estimating the types of states of the cyber systems and learn the features deeply and apply incremental learning. Two attacks detection systems are presented. One to detect attacks in completely observable cyber systems, and the other to estimate the types of states in partially observable cyber systems. Two datasets of network traffic are used for the two approaches.

In [13], two stages classifier for intrusion detection with features selection using the information gain as a ranking method is applied on UNSW-NB15 and NSL-KDD datasets.

At the first stage protocols subset is used to classify the incoming traffic flow into TCP, UDP or Other. Anomaly intrusion detection is done in second stage using different classifiers which are Decision tree, Naïve Bayes, and Artificial Neural Networks for comparison. RepTree was chosen in constructing the decision tree classifier as a fast learner using the information gain for splitting and pruning to avoid overfitting the training set. The experimental results showed that the two stage classifiers performed better in terms of speed and accuracy rate.

In [14], central Points of attribute values with apriori algorithm is applied for features selection to select the highly ranked features. Naïve Bayes and Logistic Regression are used as machine learning classifiers for intrusion detection. UNSW-NB15 and KDDCUP99 datasets was examined to evaluate the system. The pre-processing on these datasets reduced the processing time and improved the accuracy resulting in decreasing the false alarm rate.

III. PROPOSED FRAMEWORK

The objective of the proposed framework is to select the minimum number of features that achieve the highest accuracy for the classification process to detect intrusion attacks. The framework is constructed under different features selection methods. Wrapper and filter methods are selected to reach this goal where the framework can be extended to other methods such as principal component analysis and genetic algorithms. Machine learning techniques are applied for intrusion detection based on UNSW-NB15 dataset. The reasons for applying the features selection strategy are to reduce the effects of the curse of dimensionality, to minimize cost of computations and to achieve high accuracy.

A. Framework Workflow

The framework workflow as shown in figure 1, is divided into two layers. The first layer considers the different features selection methods which are used in the framework. The output from this layer consists of different sets of features. This output is forwarded to the second layer to apply the machine learning process. The best classifier under the best features selection method will be chosen after testing to represent the optimal IDS. At the first layer, the five main strategies for features selection in the proposed framework are listed in Table I. This framework is applicable for other datasets and other application domains.

TABLE I. PROPOSED FEATURES SELECTION STRATEGIES

#	Strategies
1	All features
2	Wrapper method
3	Filter using different single evaluators
4	Merged single evaluators
5	Union of the best evaluator and the best subset

The *first strategy* considers all features from the dataset and this will result in the worst performance as no features selection method is applied. It is used as the basic evaluation for comparison with other strategies. The *second strategy* is measured by applying the subset attribute evaluator using wrapper selection method, the result is a subset of features based on a machine learning classifier. Different machine learning techniques can be used to get different subsets of features for this purpose. The *third strategy* is measured by applying the filter selection method. There are different single evaluators for ranking the features under this strategy. Each evaluator is applied separately resulting in different ranking orders. The *fourth strategy* is measured after scaling the ranking orders from the different evaluators. For scaling, Eq. (1) illustrates the linear transformation for x from range $[A,B]$ to range [0,1] [15]. The average of the scaled values is calculated for each feature. A new order can be found by sorting them. Machine learning for IDS is applied based on the previous strategies. The best accuracies are calculated after testing the classifiers based on the wrapper subsets and the different evaluators. The *fifth strategy* is based on getting the union of the features from the best subset and the best evaluator.

$$f(x) = \left(\frac{\text{the variable number } (x) - \text{the min number } (A)}{\text{the max number } (B) - \text{the min number } (A)}\right) \quad (1)$$

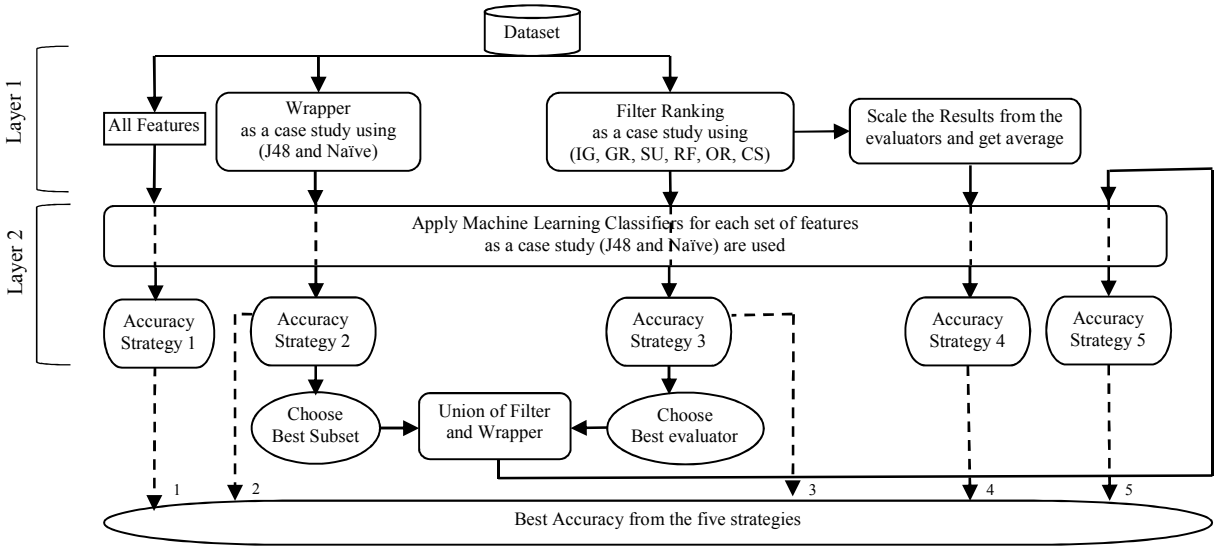

Fig. 1. Features selection Framework Workflow.

As a case study in the proposed framework, six different single attribute evaluators have been used in the proposed framework which are IG, GR, SU, RF, OR and CS. These six filters are commonly used as statistical and entropy-based methods. Two machine learning classifiers have been used which are J48 and Naïve Bayes. Each features selection method in the discussed strategies are learned and tested using these two classifiers. Other machine learning classifiers such as SVM and ANN can extend the framework as a future work.

J48 as a machine learning technique has advantages which are performing well with huge data sets and high detection accuracy rate. The disadvantage is that the construction of the decision tree is a computationally intensive task. The advantages of using the Naïve Bayes classifier are being simple and making probabilistic predictions. The disadvantage is that it can't perform regression.

At the second layer, two machine learning techniques are used which are J48 and Naïve Bayes for IDS. Training and testing are applied separately on all strategies. There are two accuracies for the first strategy one for each classifier.

In the second strategy, the subset attribute evaluator is based on training a classifier for features selection. J48 and Naïve are used for this purpose under this strategy. The output is two subsets of features one for each classifier. Two accuracies for IDS are calculated for each subset under this strategy.

In the third strategy, the filter features selection methods are applied using the six single attribute evaluators. Other filters can extend the proposed framework. The output from each filter is a ranking order for the features from the highest to the lowest relevant. For each filter the output features are pushed two by two according to their ranking order to a classifier to calculate the accuracies. The training process should stop when a saturation in accuracy is reached. As an output from this strategy there are 6 accuracies at the best number of features that gives the highest accuracy for each classifier.

In the fourth strategy, the ranking of the six filters are scaled and the average for each feature is calculated. The features are ordered in descending order according to the averages. This new order is passed to the second layer using the two classifiers. As an output there are two accuracies one for each classifier at the best number of features.

In the fifth strategy, the best features from the best evaluator and the best subset from the wrapper after testing are used as input for the union strategy. The union of features is passed to the second layer as a feedback as shown in fig. 1. The output from this strategy is two accuracies, one for each classifier.

B. Proposed framework versus related work

Table II shows a comparison between the proposed framework and some references in IDS using the UNSW-NB15 dataset as discussed in the related work. Unsupervised Correlation based features selection method has been used in reference [10]. Central points and genetic algorithm are used as features selection for IDS in references [11], [12] and [14] while the others used all features only. The proposed framework applied the wrapper and filter with different single attribute evaluators. Two new strategies are proposed, which are the merging between the six evaluators and the union of features between the best wrapper and the best filter. In comparison, the proposed framework used J48 and Naïve Bayes classifiers and can be extended with other classifiers. Some other references, as discussed in the related work, used ANN, SVM, LR and clustering techniques.

TABLE II. COMPARISON VERSUS RELATED WORK

	Proposed framework	**[7]**	**[8]**	**[9]**	**[10]**	**[11]**	**[12]**	**[13]**	**[14]**
Features selection	Wrapper Filter (IG, GR, SU, RF, OR, CS)	×	×	×	Unsupervised Correlation	Central Points ARM	Genetic algorithm	IG	Central Points with apriori algorithm
Union between wrapper and filter	√	×	×	×	×	×	×	×	×
Merge the six evaluators	√	×	×	×	×	×	×	×	×
Applied machine learning classifiers	J48 Naïve Bayes	J48 Naïve Bayes LR SVM	SVM	J48 ANN	ANN	Naïve Bayes Clustering LR	CNN K-means	J48 Naïve Bayes ANN	Naïve Bayes LR

IV. EXPERIMENTAL RESULTS

This section presents the results and the evaluation of the different classifiers under each strategy for features selection using the proposed framework. UNSW-NB15 [16] has been used as one of the recent datasets in intrusion detection. It consists of 42 features and one label attribute that marks normal or an attack. The features are listed in table III.

TABLE III. THE FEATURES OF THE UNSW-NB15 DATASET

#	Name	#	Name
1	dur	22	Dtcpb
2	proto	23	dwin
3	service	24	tcprtt
4	state	25	synack
5	spkts	26	ackdat
6	dpkts	27	smean
7	sbytes	28	dmean
8	dbytes	29	trans_depth
9	rate	30	response_body_len
10	sttl	31	ct_srv_src
11	dttl	32	ct_state_ttl
12	sload	33	ct_dst_ltm
13	dload	34	ct_src_dport_ltm
14	sloss	35	ct_dst_sport_ltm
15	dloss	36	ct_dst_src_ltm
16	sinpkt	37	is_ftp_login
17	dinpkt	38	ct_ftp_cmd
18	sjit	39	ct_flw_http_mthd
19	djit	40	ct_src_ltm
20	swin	41	ct_srv_dst
21	stcpb	42	is_sm_ips_ports

The training set is 175,341 records, the normal percentage is 32% and anomaly is 68%. The number of records in the testing set is 82,332 records, the normal percentage is 45% and anomaly is 55%.

The proposed framework is implemented using Weka software [17] to evaluate the accuracy and performance of the different aspects. The specs of the testing machine are Intel i5 quad Core Processor, @ 2.7 GHz, 6 MB Cache, 4GB RAM, Windows 7 operating system.

Table IV presents the accuracy measures for the proposed framework grouped by each features selection strategy. The best number of features that gives the highest accuracy in each case is mentioned. The accuracy measures are divided into two columns, one for each classifier. The accuracy is measured as in Eq. (2).

$$Acc = \frac{TP+TN}{TP+TN+FP+FN} \quad (2)$$

The results show that the best classifier with the highest accuracy is by using J48 with the GR filter at 18 features which achieves 88.3%. The list of these 18 features according to Table III is {10, 11, 32, 42, 4, 26, 24, 25, 17, 13, 8, 6, 9, 7, 28, 1, 35, 30}. Table V shows the confusion matrix under this classifier by listing the percentage of instances in the four categories for evaluating the classification process.

Figure 2 presents the effect of increasing the acceptable number of features according to the ranking order of the GR filter using J48. It is apparent that the accuracy is increasing by accepting more features until reaching saturation. At this point, 18 features are enough for the classification process. Under this strategy a speedup factor of 2 can be gained comparing versus the first strategy using all features.

TABLE IV. THE BEST ACCURACY AND NUMBER OF FEATURES FOR THE PROPOSED FRAMEWORK

	#	Strategies		UNSW-NB15	
				J48	Naïve
Strategy 1	1	All features	# features	42	42
			Accuracy	87%	76.9%
Strategy 2	2	Wrapper with J48	# features	24	24
			Accuracy	86.6%	0.764
	3	Wrapper with Naïve	# features	5	5
			Accuracy	84.2%	80.7%
Strategy 3	4	Filter CS	# features	24	28
			Accuracy	88.1%	77.9%
	5	Filter GR	# features	18	14
			Accuracy	88.3%	78.5%
	6	Filter IG	# features	24	28
			Accuracy	87.8%	78%
	7	Filter OR	# features	30	32
			Accuracy	87.6%	79.4%
	8	Filter RF	# features	42	14
			Accuracy	87.0%	78.3%
	9	Filter SU	# features	28	6
			Accuracy	88.2%	80.1%
Strategy 4	10	Union	# features	33	10
			Accuracy	86.8%	76.7%
Strategy 5	11	Merge	# features	26	30
			Accuracy	88.2%	78.6%

TABLE V. THE CONFUSION MATRIX FOR THE BEST CLASSIFIER

		Predicted	
		Pass (Positive)	Fail (Negative)
Actual	Pass (True)	TP = 35%	FN = 10%
	Fail (False)	FP = 2%	TN = 53%

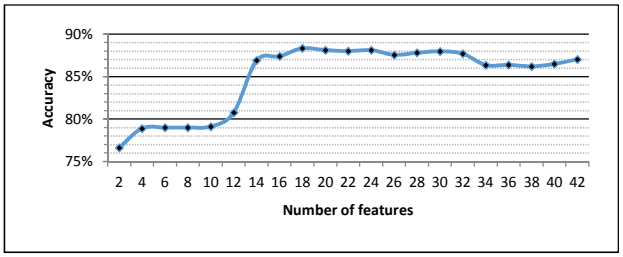

Fig. 2. The accuracy of the GR evaluator using J48.

The minimum number of features can be extracted using the wrapper method that achieves 5 features with accuracy of 84.2% using Naïve Bayes for features selection and J48 for intrusion detection. This strategy can give a speedup factor of 4 and a loss of accuracy of around 3% from the best accuracy measure.

The proposed framework introduced the merged evaluators' strategy and the union between the best wrapper subset and the best evaluator. No more enhancement in accuracy has been achieved under the used dataset. However, these strategies can be tested on other datasets and applicable for other applications.

The proposed framework is an offline process to decide how to deal with online traffic for intrusion detection. Table VI shows the average execution time for intrusion detection using the whole test set under the different strategies from 10 experimental readings. All features strategy with J48 needs 1.578 sec. GR with J48 needs 0.79 sec and wrapper with Naïve Bayes needs 0.382 sec where a speedup factor of 2 and 4 can be gained respectively over the first strategy. The average testing time for one record is 9.5953 microseconds using the GR at 18 features with J48 as the best classifier.

TABLE VI. THE EXECUTION TIME FOR THE DIFFERENT STRATEGIES

	#	Strategies	Average time (Seconds)	
			J48	Naïve
Strategy 1	1	All features	1.578	3.367
Strategy 2	2	Wrapper with J48	0.933	1.869
	3	Wrapper with Naïve	0.382	0.622
Strategy 3	4	Filter CS	1.086	2.495
	5	Filter GR	0.79	1.123
	6	Filter IG	1.055	2.073
	7	Filter OR	1.379	2.491
	8	Filter RF	1.521	1.312
	9	Filter SU	1.12	0.68
Strategy 4	10	Union	1.132	0.936
Strategy 5	11	Merge	1.071	2.28

V. CONCLUSION AND FUTURE WORK

This paper presents a framework that applies different strategies for features selection using filter and wrapper methods. J48 and Naïve Bayes algorithms are used as classifiers on the UNSW-NB15 dataset. The experimental results show that, the best strategy is by using 18 features from the GR ranking method and applying J48 as a classifier getting an accuracy of 88% and a speedup factor of 2.

As a future work SVM and ANN will be used to extend the framework under the different features selection strategies. In addition, majority voting scheme between all classifiers can be used to increase the accuracy in IDS. A parallel model can be designed for this purpose where the problem is parallel by nature. The same framework is currently being applied on other datasets namely NSL-KDD.

REFERENCES

[1] SANS Institute InfoSec Reading Room, "Understanding Intrusion Detection Systems", Available: https://www.sans.org/reading-room/whitepapers/detection/understanding-intrusion-detection-systems-337, [Accessed: November 2017].

[2] Sailesh Kumar, "Survey of Current Network Intrusion Detection Techniques", Available: http://www.cse.wustl.edu/~jain/cse571-07/ftp/ids/, [Accessed: November 2017].

[3] Jean Philippe Planquart, "Application of Neural Networks to Intrusion Detection", SANS Institute InfoSec Reading Room.

[4] Ian H. Witten and Eibe Frank, "Data Mining: Practical Machine Learning Tools and Techniques", Morgan Kaufmann Publishers, Publication Date: January 20, (2011) | ISBN-10: 0123748569 | ISBN-13: 978-0123748560 | Edition: 3

[5] Binita Kumari andTripti Swarnkar, "Filter versus Wrapper Feature Subset Selection in Large Dimensionality Micro array: A Review", International Journal of Computer Science and Information Technologies (IJCSIT), Vol. 2 (3) , 2011, pp.1048-1053

[6] Jasmina Novaković, Perica Strbac, Dusan Bulatović, "Toward optimal feature selection using Ranking methods and classification Algorithms", Yugoslav Journal of Operations Research, Vol. 21, No. 1, pp.119-135, 2016.

[7] Deval Bhamare, Tara Salman, Mohammed Samaka, Aiman Erbad and Raj Jain, "Feasibility of Supervised Machine Learning for Cloud Security", in Procedding of the 3rd International Conference on Information Science and Security (ICISS2016), December 19th - 22nd, 2016, Pattaya, Thailand.

[8] Amar Agrawal, Sabah mohammed and Jinan Fiaidhi, "Developing Data Mining Techniques for Intruder Detection in Network Traffic", International Journal of Security and its Applications, Vol. 10, No.8, pp.335-342, 2016.

[9] Lennart van Efferen, "Adaptive Cyber Security: Machine Learning Techniques for a Flow-Based Anomaly Intrusion Detection System", Bachelor Thesis, Leiden Institute of Advanced Computer Science (LIACS), Leiden Universit, 2017.

[10] Mohamed Idhammad, Karim Afdel and Mustapha Belouch, "DoS Detection Method based on Artificial Neural Networks", International Journal of Advanced Computer Science and Applications (IJACSA), Vol. 8, No. 4, pp. 465-472, 2017.

[11] Nour Moustafa and Jill Slay, "A Hybrid Feature Selection for Network Intrusion Detection Systems: Central Points and Association Rules", in Procedding of the 16th Australian Information Warfare Conference (pp. 5-13), held on the 30 November - 2 December 2015, Edith Cowan University, Joondalup Campus, Perth, Western Australia.

[12] Sayantan Guha, "Attack Detection for Cyber Systems and Probabilistic State Estimation in Partially Observable Cyber Environments", Master of Science Thesis, Arizona State University, 2016.

[13] Mustapha Belouch, Salah El Hadaj, and Mohamed Idhammad, "A Two-Stage Classifier Approach using RepTree Algorithm for Network Intrusion Detection", International Journal of Advanced Computer Science and Applications (IJACSA), Vol. 8, No. 6, pp. 389-394, 2017.

[14] Dipali Gangadhar Mogal, Sheshnarayan R. Ghungrad and Bapusaheb B. Bhusare, "NIDS using Machine Learning Classifiers on UNSW-NB15 and KDDCUP99 Datasets", International Journal of Advanced Research in Computer and Communication Engineering (IJARCCE), Vol. 6, Issue 4, pp. 533-537, 2017.

[15] Ajitesh Kumar, "How to Scale or Normalize Numeric Data using R", Available: https://vitalflux.com/data-science-scale-normalize-numeric-data-using-r/, [Accessed: November 2017].

[16] "UNSW-NB15 dataset", Available: https://cloudstor.aarnet.edu.au/plus/index.php/s/2DhnLGDdEECo4ys?path=%2F, [Accessed: November 2017].

[17] "Waikato environment for knowledge analysis (weka) version 3.7.10.", Available: http://www.cs.waikato.ac.nz/ml/weka/, [Accessed: November 2017].