FACTA UNIVERSITATIS
Series: **Electronics and Energetics** Vol. 32, N° 2, June 2019, pp. 315-330
https://doi.org/10.2298/FUEE1902315O

FEATURE SELECTION FOR INTRUSION DETECTION SYSTEM IN A CLUSTER-BASED HETEROGENEOUS WIRELESS SENSOR NETWORK

Opeyemi Osanaiye[1], Olayinka Ogundile[2], Folayo Aina[3], Ayodele Periola[4]

[1]Department of Telecommunication Engineering, Federal University of Technology, Minna, Niger State, Nigeria
[2]Department of Physics and Telecommunications, Tai Solarin University of Education, Ogun State, Nigeria
[3]Department of Telecommunication Science, University of Ilorin, Ilorin, Kwara State, Nigeria
[4]Electrical Electronics and Computer Engineering, Bells University of Technology, Ota, Nigeria

Abstract. *Wireless sensor network (WSN) has become one of the most promising networking solutions with exciting new applications for the near future. Notwithstanding the resource constrain of WSNs, it has continued to enjoy widespread deployment. Security in WSN, however, remains an ongoing research trend as the deployed sensor nodes (SNs) are susceptible to various security challenges due to its architecture, hostile deployment environment and insecure routing protocols. In this work, we propose a feature selection method by combining three filter methods; Gain ratio, Chi-squared and ReliefF (triple-filter) in a cluster-based heterogeneous WSN prior to classification. This will increase the classification accuracy and reduce system complexity by extracting 14 important features from the 41 original features in the dataset. An intrusion detection benchmark dataset, NSL-KDD, is used for performance evaluation by considering detection rate, accuracy and the false alarm rate. Results obtained show that our proposed method can effectively reduce the number of features with a high classification accuracy and detection rate in comparison with other filter methods. In addition, this proposed feature selection method tends to reduce the total energy consumed by SNs during intrusion detection as compared with other filter selection methods, thereby extending the network lifetime and functionality for a reasonable period.*

Key words: *Chi-squared, cluster, Gain ratio, intrusion detection, NSL-KDD, ReliefF, WSNs*

Received January 16, 2019; received in revised form March 9, 2019
Corresponding author: Opeyemi Osanaiy
Department of Telecommunication Engineering, Federal University of Technology, Minna, Niger State, Nigeria
(E-mail: opyosa001@myuct.ac.za)

1. INTRODUCTION

Wireless sensor networks (WSNs) are formed by sets of distributed autonomous devices with the capability to sense, process, transmit and receive observed or measured condition. The sensor nodes (SNs) used in WSNs are characterized by their light weight, limited processing power, limited energy, low storage capacity, short communication range and low bandwidth [1]. The sensor component of the SN measures the observed condition of a particular situation or physical surroundings while the microprocessor ensures the obtained information are intelligently computed [2]. The wireless radio of the node, on the other hand, ensure communication between neighbouring nodes.

WSNs often times are deployed in remote, harsh and unattended environment over a certain period of time. These locations are most times not accessible, therefore, it is impractical to carryout maintenance on the nodes after installation. Common among its applications are in the area of environmental monitoring, air craft control, disaster control, medical health monitoring, surveillance and military application among many others [3].

Although WSNs have been used in numerous applications, the requirements of these applications have put a lot of constraints on its design and deployment. Security has been identified in the literature as one of the main constraints in the deployment of WSNs. This is evident as WSNs are subjected to vulnerabilities associated with wireless communication. Additionally, in events that involve unprotected hostile outdoor environment, WSNs are prone to different types of attack that compromise the confidentiality, integrity, authentication and availability of the data traffic and battery life of the SNs [4,5]. Many of these attacks have been identified, analysed and discussed in the literature, with authors proffering different defence and prevention techniques. One of such attacks is the denial of service attack, which can also be referred to as packet drop attack or sinkhole attack [6]. Blackhole attack in WSN is also a type of denial of service attack that advertises itself as either the destination node or the shortest route to get to the destination. Upon receiving these falsely advertised packets from other nodes, the attacker discards all the packets. Selective forwarding is a derivative of blackhole attack in which the adversary node does not reject all received packet, instead, it randomly selects packets that will be discarded [7]. The adversary can use this to evade detection.

In order to protect the WSNs from intrusion by an adversary, various intrusion detection system (IDSs) have been proposed by researchers. These IDS defence solutions are categorized into signature-based and anomaly-based. The former relies on signatures of known attack patterns while the latter profiles a statistical usage model over a certain amount of time to classify data packets as either normal or anomaly using various techniques such as data mining, machine learning and statistical modelling. The signature-based has a major flaw of not being able to detect unknown attacks while anomaly-based detection suffer from high false positive rate [8]. This has necessitated the emergence of a hybrid solution that uses the complementary feature of both techniques to achieve a higher detection rate. The novel challenges of most of these proffered security solutions for WSNs include its limited storage capacity, computational resources and battery power. Therefore, traditional security solutions are inappropriate for WSNs.

Due to the resource limitation in WSN environment, proposed IDS designs are often lightweight and highly specialized by type of attack to reduce false alarms. Computational Intelligence IDS improves its performance by providing features such as learning, reasoning, perception, evolution and adaptation [5]. These features can be explored to

develop a more robust IDS that is adaptive to different application scenarios, to handle unknown attacks.

In this work, we introduce a pre-processing phase in the form of feature selection by combining three filter feature selection methods; gain ratio, chi-squared and ReliefF, herein called triple-filter, to select one-third split (14 most important features) from the original dataset before classifying with a decision tree algorithm. The motivation behind feature selection is the resource constraint of SNs, therefore machine learning techniques use feature selection to eliminate redundant features to reduce the complexity of the proposed system. Intrusion detection benchmark dataset, NSL-KDD that consists of 41 features [9] was used to evaluate the performance of the IDS by considering the detection rate, classification accuracy and false alarm rate in Waikato environment for knowledge analysis (Weka). Furthermore, we compared our result with the proposed work in [10]. The results obtained show that our proposed method can effectively reduce the number of features with a high classification accuracy and detection rate and a low false alarm rate as compared with [10].

The contribution and relevance of this paper is as follows. In this work, we introduce a pre-processing phase in the form of feature selection, similar to our approach in [11]. However, here we combine three filter feature selection methods, herein called triple-filter. This is used to select the 14 most important features in NSL-KDD for intrusion detection in WSNs. This reduces the complexity of the IDS by presenting a lightweight technique. Reduced IDS complexity implies that the SNs in a WSN will consume less energy while maintaining high availability. Since the SNs are battery powered, prolonging the network lifetime and functionality to a reasonable time is very paramount. Thus, our proposed IDS defence solution is suitable for use in a real-time WSN as it helps to efficiently extend the network life-time and functionality.

The rest of the paper is structured as follows. Section 2 describes related work on IDS defence solution for WSNs. In Section 3, the WSN architecture and the proposed IDS was discussed. Also, the section explains the three filter feature selection methods; gain ratio, chi-squared, and RefliefF in details. The feature selection and execution process is highlighted in Section 4 while Section 5 present the experimental results. Section 6 highlights the performance measure with respect to the classification accuracy, detection rate and false alarm, while we discuss the results in Section 7. Finally, Section 8 concludes the work and suggests possible research directions.

2. RELATED WORK

In defending against malicious attacks in WSN, various intrusion detection approaches have been proposed in the literature. An intelligent intrusion and prevention system was proposed in [1] by introducing a specialized dataset for WSN. This improves the detection and classification of four types of denial of service (DoS) attacks: Blackhole, Grayhole, Flooding, and Scheduling attacks. Artificial Neural Network (ANN) was used to train the dataset to detect and classify the different DoS attacks. Results from the work show that the dataset, WSN-DS, enhanced the IDS ability to achieve a higher classification accuracy rate. An IDS based on evidence theory was proposed in [12] for cluster-based WSN. In this work, each cluster head collects the behavioural pattern of its cluster members before constructing an input evidence according to the deviation from the normal pattern. A weight value is further

applied to represent the importance of each behaviour characteristics and revise the evidence before its synthesis. A hybrid IDS that enhances security in cluster-based WSN has been proposed in [13]. In this work, the proposed IDS is deployed on the cluster head and consists of both an anomaly and misuse module. The output of the anomaly and misuse modules are integrated with a decision-making module to identify the presence of an attack before subsequently classifying into different attack type. In [14], a distributed two-layer and three-layer IDS scheme was proposed for WSN to detect intrusion using 10% of the data to learn during the training phase. A complexity reduction process was introduced to select the features to minimize the energy consumption.

A specification-based intrusion detection system was proposed in [15]. This system uses rule-based technique to map behaviours to either normal or anomalous. The rule-based technique optimizes the local information obtained by watch dogs into a global information for decision making by cluster heads. This compensate for the communication pattern in the network. In [16], a decentralized IDS was proposed for WSN. The proposed algorithm is divided into three phases; data acquisition, rule application and intrusion detection. In data acquisition phase, messages are obtained in promiscuous mode and the relevant information are filtered and subsequently stored for analysis. The rule application phase, on the other hand, process the information and apply the rule to the stored data. If the message fails the test during analysis, a failure is raised. Lastly, in the intrusion detection phase, the amount of raised failure is compared with the expected amount of occasional failures in the network. Intrusion alarm is raised if the former is higher than the latter.

In [10], an integrated intrusion detection system (IIDS) was proposed for cluster-based WSNs. The IIDS was based on an earlier work in [17] and it consists of three individual IDSs, namely: intelligent hybrid intrusion detection system (IHIDS), hybrid intrusion detection system (HIDS) and misuse IDS. These IDSs are designed for the base station (BS), cluster head and cluster members, based on their capacity and the type of attack they are vulnerable to. For example, the IHIDS with a learning capability is deployed in the BS. The IHIDS combines the anomaly and misuse detection by first filtering a large number of normal packets. The packets are then forwarded to the misuse detection module to identify the type of attack. This is done to achieve a high detection with low false alarm. The cluster heads, on the other hand, houses the HIDS, which is similar to the IHIDS but without a learning ability. The HIDS function to optimally detect attacks, however, it retrains the behaviour of the new attack previously detected and classified by the IHIDS. Lastly, due to the resource constraints of SNs, the misuse IDS is proposed. The misuse IDS uses a predetermined attack model to match packets to find and detect attacks. Experimental results for the misuse detection, using back propagation network and KDDCup' 99 dataset, shows that a detection rate of 90.96% was achieved with an accuracy of 99.75% and false positive of 2.06%.

In the discussion above, different techniques have been considered for feature selection in WSNs. The overall aim of these techniques is to enhance the ability of SNs to differentiate attacks in WSN. The performance of the security mechanism designed in this manner can be influenced by the number of features of the dataset. Different kinds of feature selection methods can be used to achieve varying results. This is because of the resource limitation characterizing WSNs. Therefore, a combination of different feature selection methods that considers the resource constraints in WSN is required. A strategy that uses multiple algorithms that harness its features will be advantageous in classifying the type of attack in WSNs.

Considering the resource limitation that characterize WSNs, this work proposes a feature selection method by combining the trio of gain ratio, chi-squared and ReliefF (triple-filter) to select one-third split (14 features) from the initial 41 features of the dataset. This will significantly reduce the complexity of the IDS and minimize the energy consumed during intrusion detection. More so, this filter feature selection method offers high detection rate with good classification accuracy and a low false alarm rate as shown in Table 4.

3. WSN ARCHITECTURE AND PROPOSED IDS

The deployment of WSNs is often made up of tens to hundred thousand of autonomous SNs that function via member node communication. This is necessary as a single sensor node only covers a small area, therefore can only provide limited information. This single node deployment limitation has brought forth the introduction of networks of SNs, that are self-organising and collaborative, to achieve a wider coverage over a large environment. The SNs monitor, sense, computes and transmits the observed and measured condition of the environment to relay the information to the intended user through the base station. A typical sensor node consists of sensor components, microprocessor components and wireless radio. The sensor component measure the condition of the observed environment of interest while the microprocessor component embedded in the node is used to intelligently compute the obtained information [1]. The wireless radio component of the sensor node is used to initiate communication between neighbouring sensor nodes in WSN. A significant benefit of sensor network deployment is its ability to extend its coverage area to environments where it is near impossible for human beings to access.

When categorizing WSNs, environment the sensor nodes are deployed can be used. The work in [18] described five types of WSN, namely: underground WSN, terrestrial WSN, underwater WSN, multi-media WSN and mobile WSN.

In underground WSN deployment, sensor nodes are buried under the surface of the ground to monitor and sense its condition. These sensor nodes transmit the sensed information to the sink node, which is placed above the ground, to relay it to the base station. Terrestrial WSN, on the other hand, consist of several cheap sensor nodes deployed on a specific area of interest, on the surface of the earth in a pre-planned or ad hoc way. The pre-planned deployment involves the optimal placement of sensor nodes, such as grid placement and 3-D placement model [19], while in ad hoc deployment, sensor nodes are randomly deployed. Underwater WSN deployment are instances where the sensor nodes are deployed under the water body to sense, explore and gather information about a subject matter and transmits this information using acoustic wave [20]. Underwater WSN presents a sparse sensor node deployment as compared to the dense deployment of terrestrial WSN. Multi-media WSN are sensor nodes equipped with cameras and microphones to ensure the efficient monitoring and tracking of multi-media events, such as imaging, audio and video [21]. Here, the sensor nodes interconnects over a wireless medium to retrieve, process, compress and convey sensed data in a pre-planned arrangement to ensure coverage. One major obstacle to the deployment of multimedia WSN is the resource challenge of sensor nodes, due to the excessive energy consumption during the compression and decompression when transmitting multi-media events. Finally, mobile WSN are sets of sensor nodes deployment that move and interact with the physical environment. Just as with static WSN,

mobile nodes can sense, compute, transmit and receive observed and measured events. The sensor nodes have the potential to reorganise and reposition themselves after deployment to obtain information. The obtained information can be distributed among other mobile nodes within their communication range using dynamic routing protocol.

WSN can be further classified according to the structure and uniformity of the deployed sensor nodes. Some deployment consists of uniform nodes with equal capacity while other deployments consist of different sizes and capacity, depending on the architecture. In WSN, the network structure (topology) can be categorized into two, namely: flat-based and hierarchical [22]. The flat-based topology consists of sensor nodes with equal capacity, playing similar roles, such as monitoring and sensing events, computing the sensed information and transmitting it directly or via multi-hop routing towards the BS [23]. On the other hand, hierarchical WSNs are designed to distribute the sensing and monitoring function of the SNs into different levels. Cluster-based WSNs are typical example of hierarchical WSNs. In this paper, we limit our scope to cluster-based WSNs.

Arranging SNs into clusters have been widely employed by researchers to efficiently sense and monitor a particular environment. The clustering technique is widely used in WSNs because it offers advantages such as reduced energy consumption, fault-tolerance, scalability, efficient data aggregation, latency reduction, and robustness [3,24]. A clustered WSN comprises of two sets of nodes, namely: the member nodes known as the non-cluster head nodes, and the coordinating nodes often referred to as the cluster head. Fig.1 shows a typical example of a cluster-based WSN, where c represents a cluster. As shown in Fig. 1, the non-cluster head nodes forward the sensed message to their respective cluster heads in a process known as intra-cluster communication. The cluster heads organise the messages from their respective members before transmitting it to the BS. Thus, clustering technique can be regarded as a two-layer hierarchy WSN, where the cluster heads work in the upper layer and the non-cluster head nodes operate in the lower layer. The coordinating nodes in most cases perform more function as compared to the lower layer nodes. Therefore, the cluster head nodes are sometimes equipped with better processing subsystem, sensing unit, radio subsystem, and power supply unit as compared with the lower layer. If the components of all the sensor nodes in the network are the same, the clustering WSN is usually referred to as a homogeneous clustering WSN. Otherwise, it is referred to as a heterogeneous clustering WSN.

In this work, we assume that the cluster heads are equipped with a better processing subsystem, sensing unit, radio subsystem, and power supply unit. Accordingly, our proposed IDS is deployed on the cluster heads for intrusion detection. The cluster heads will monitor the SNs to detect attacks. Furthermore, the cluster heads will filter abnormal data and forward all the reliable sensed information to the BS, either directly or via one or more relay nodes. From the literature, the relay nodes can either be a cluster head node or a non-cluster head node [3]. Since our proposed IDS are installed only on the cluster head nodes, we assume that the relay nodes towards the BS can only be a cluster head node in order to maintain high availability. More so, the IDS is deployed only on the cluster head nodes to conserve the battery energy of the non-cluster head nodes, which in turns prolong the network lifetime and functionality.

Finally, the BS integrates all the collected information and transmits the final result to the end user. This proposed IDS defence solution can be deployed with relevant energy-efficient and energy-balanced clustering routing protocols such as [25, 26, 27, 28, 29]. However, in this paper, we verify our proposed IDS solution with the routing algorithm proposed in [26].

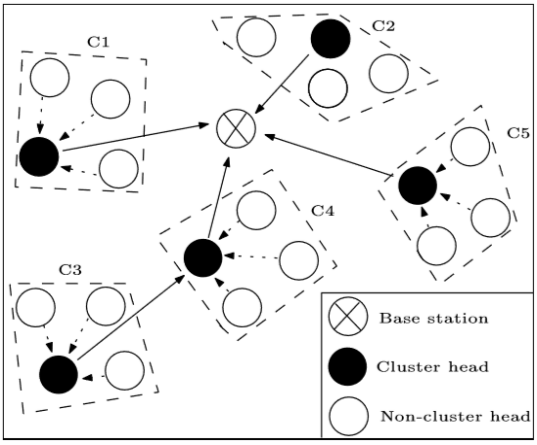

Fig. 1 Typical example of a cluster-based WSN

In this section, we present a detailed explanation of our proposed IDS. Current feature selection methods can be categorised into filter, wrapper and embedded method. While wrapper and embedded methods are time consuming and require specific classification techniques to determine the importance of feature subset, filter methods often rely on the general attributes of the dataset to carry out data pre-processing; a step which is independent of the induction algorithm [11]. Furthermore, filter methods can be classified into univariate and multivariate techniques. Univariate techniques, such as information gain, presents an efficient and scalable method, however they tend to disregard feature dependencies. Multivariate filter techniques, on the other hand, incorporate feature dependencies. This makes multivariate techniques more complex. The system that use multivariate techniques are less scalable and have a longer computational time than systems incorporating univariate techniques.

In this work, we combine three filter selection methods, gain ratio, chi-squared, ReliefF, herein referred to as triple-filter method. The choice of these filter methods is due to its ranking and space searching algorithm. Furthermore, research has shown that combining feature selection methods can improve the performance of classifiers by identifying features that are weak individually but strong as a group [31]. Our proposed triple-filter method relies on the combined strength of the trio to determine the features that are strong in determining the output class. Here, we select 14 most important features.

A. Gain Ratio

In filter feature selection, the value of gain ratio is said to be large when data are evenly spread while it presents a small value when all data belongs to only one branch of attribute. Gain ratio is an improvement on information gain that remedies its bias towards features with large diversity value exhibited by the latter. It uses the number and size of branches to determine an attribute and corrects information gain by using intrinsic information [30]. Intrinsic information is the entropy of distribution of instance value for a given feature. Gain ratio can be calculated [30] for a given feature x and a feature value of y using the equation (1) below

$$Gain\ Ratio\ (y,x) = \frac{Information\ Gain(y,x)}{Intrinsic\ Value(x)} \quad (1)$$

Where, Intrinsic Value $(x) = -\sum \frac{|S_i|}{|S|} * Log_2 \frac{|S_i|}{|S|}$

$|S|$ is the number of possible values feature x can take while $|S_i|$ is the number of actual values of feature x. In our work, we select 14 features from the NSL-KDD dataset that represents the highest ranked feature using gain ratio.

B. Chi-Squared

Chi-squared (χ^2), in mathematical statistics, is a feature selection method that is often used to determine the worth of an attribute with respect to a particular class. Chi-squared can be used to test the independence of two variables with an initial hypothesis, H_0 ; with the assumption that the two features are not related [30, 31]. This can be tested using the chi-squared formula:

$$\chi^2 = \sum_{u=1}^{r} \sum_{v=1}^{c} \frac{(A_{u,v} - P_{u,v})}{P_{u.v}} \quad (2)$$

where $A_{u,v}$ is the actual value and $P_{u,v}$ is the predicted value declared by the hypothesis H_0. The higher the value of the chi-squared, the higher the evidence against the null hypothesis.

C. ReliefF

ReliefF is an extension of an earlier Relief algorithm that randomly samples an instance from the dataset to locate its nearest neighbours from both the same and opposite class [32]. The values of the attributes obtained from the nearest neighbours, after comparing with the sample instance, are used to update the relevant score from each of the attributes. The idea behind this is that, significant attributes will be able to distinguish between instances that belong to different classes and have the same value from instances belonging to the same class [32]. Key among the advantage of ReliefF filter method is its ability to deal with multiclass issues and its robustness and ability to deal with noisy and incomplete data [33]. ReliefF can be applied in virtually all situations because of its low bias.

4. Feature Selection and Execution Process

As depicted in Fig. 2, we divided our proposed IDS defence solution for cluster-based heterogeneous WSN into three phases. The first phase in implementing a lightweight IDS is to introduce an initial pre-process stage for the dataset prior to training. To achieve this, we use our proposed triple-filter method for ranking. By ranking, the features that are strong in determining the output class of the dataset are obtained and one-third split of the ranked features are selected (that is, 14 features). One-third features of the entire features in the dataset was arrived at after ranking and eliminating redundant features before the performance of the classifier start to decline. These features selected represents the most significant features among all the filter methods. In the second phase, the training phase, the features selected after pre-processing the NSL-KDD dataset are used to train the IDS to detect possible attacks in the network. This is deployed on the cluster

head to monitor data from the sensor nodes to the base station. The final phase, the classification phase, is a process whereby a labelled training dataset is used to learn, before subsequently classifying a test data into one of the class labels [34].

Anomaly detection techniques that use classification-based algorithms can be divided into two stages; the training stage and testing stage. In the training stage, labelled data are used to learn a particular classifier. Subsequently, this classifier can be used in the test stage to classify a test instance as either normal or anomaly. In this work, we use decision tree classification algorithm to detect the occurrence of a DoS attack.

Decision trees are data mining approach which are often called classifier trees or hierarchical classifiers and are used for prediction. It is a popular method because of its simple structure, ease to understand and the short time required to interpret [35]. During the classification process, the degree of adjustment of the model to the training set is very essential. When a tight stopping criterion is employed, it often creates a small and unidentified decision tree, while the algorithm with a loose stopping criterion, on the other hand, gives a larger decision tree that tends to over-fit the training dataset.

Decision tree has been embraced for classification and data analysis in fields, such as agriculture, environmental, health, etc. Decision trees are recursive partition models that use a single variable to divide datasets at each level. Initially, all sets of cases are defined to belong to the same class before a variable is selected, using a split criterion, to determine the attribute to insert in a node and branch. Decision tree nodes consist of set of rules where each tree node is labelled with an attribute variable which creates a branch for each value. They are represented by a tree like structure, with the leaf nodes labelled with a class label [36].

From its original version of ID3 (Iterative Dichotomiser 3), C4.5 and C5.0 has been developed as an advance version of ID3 [35]. Over the years, the C4.5 algorithm has been used in the literature as the standard model for supervised learning. During a classification process, a training dataset is used to train the decision tree algorithm while a test dataset is used to validate the model. When there is a new sample of a test dataset, a prediction can be made on the state of the class variable using the path of the tree from the root to the leaf node, for the tree structure and sample values.

For example, let us consider a set S, and select a case at random belonging to class C_t. To determine if the random sample belongs to the class C_t, we find P_i using the equation [37]:

$$P_i = \frac{freq\ (C_t, S)}{|S|} \tag{3}$$

Where $t = \begin{cases} 0, & normal \\ 1, & attack \end{cases}$

$|S|$ denotes the number of samples contained in the set S. The information conveyed can therefore be represented by $-log_2 P_i$, $P_i \epsilon P$; $P = \{P_1, P_2, P_3 \ldots\ldots\ldots\ldots, P_n\}$ where P is the probability distribution. The entropy of P, which is the information conveyed by the distribution, can be expressed as follows:

$$Info\ (P) = \sum_{i=1}^{n} -P_i\ log_2 P_i, \tag{4}$$

where *n* is the length of the information. When a set of K samples are segmented by using a non-categorical attribute X, we have a set $K = \{K_1, K_2, K_3 \ldots, K_m\}$ where m is the number of samples. The weighted average is the information used in determining the class of an element K_i and can be determined using the formula:

$$Info\ (X,K) = \sum_{i=1}^{m} \frac{|K_i|}{K} \times\ Info(K_i) \tag{5}$$

Therefore, the information gain can be computed as follows:

$$Gain\ (X,K) = Info\ (K) - Info\ (X,K) \tag{6}$$

The Eqn. 6 above expresses the difference between the information required to determine the value of an element of K and the information required to determine K, having obtained the value of the attribute X. This is therefore referred as the information gain due to attribute X. In this work, we use *J48* decision tree classification algorithm, a version of the *C4.5* for classification.

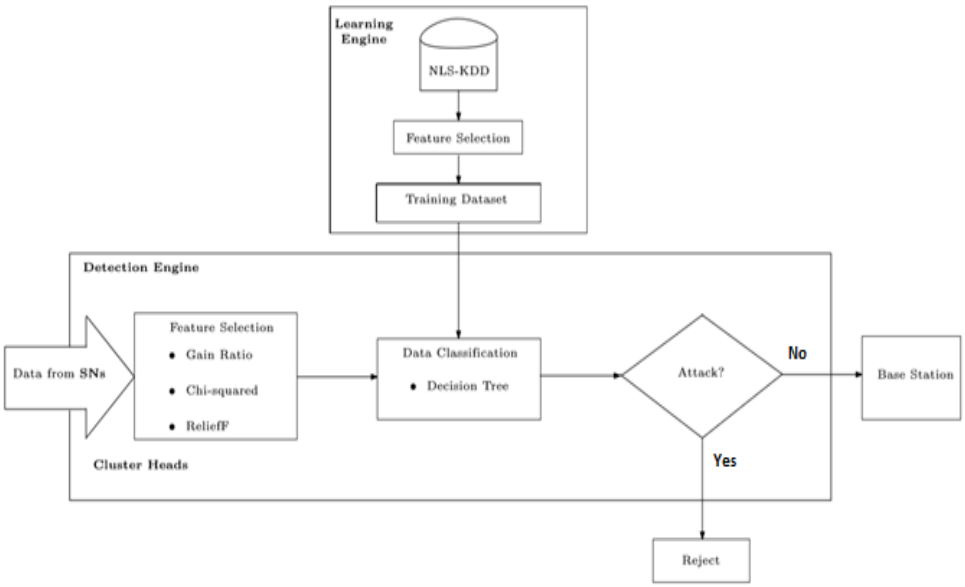

Fig. 2 Proposed Intrusion Detection Model

5. EXPERIMENTAL RESULTS

In this work, we use the combination of three filter methods during the pre-processing stage to select features from the labelled dataset, NSL-KDD. The most relevant features that are strong in determining the output class are ranked and chosen to be used by the machine learning algorithm to classify traffic packets as either normal or anomaly. Weka software [38], a machine learning tool that consists of series of machine learning algorithms, is used for our experimental analysis. During classification, the parameters of Weka are set to its default values.

During evaluation, we determine the performance of our proposed triple-filter method by using an open source NSL-KDD dataset. The motivation behind the use of NSL-KDD in our work is because it is open source and readily available online. Furthermore, NSL-KDD can be modified to suit different experimental attack scenario in WSN. The NSL-KDD is a labelled benchmark dataset developed from the initial KDDCUP'99 that presented some

shortcomings. The NSL-KDD consists of 41 features and 2 classes, labelled as either attack or normal. The features in the dataset are categorized into four groups, namely; basic features, content features, time-based traffic features and connection-based traffic features [9]. The attacks in the dataset are grouped into DoS, Probe, R2L and U2R; with these attacks divided into training set and test set. The training set consists of 21 attack types, while on the other hand, the test set consists of an extra 17 unique attack type [9]. In this work, we have modified the dataset and extracted the DoS attack trace. DoS attack is one of the most prevalent attacks on the resource constraint sensor nodes in WSN that depletes its energy and cause a denial of service. DoS attacks on systems, often a times, are carried out using similar methods, however, its impact on different hosts varies.

The feature selection process is carried out to determine the one-third slip (14 highest ranked features) of the NSL-KDD dataset using our proposed triple filter method, as shown in Table 1. This experiment is performed on an HP 64*bit* Windows 10 operating system with Intel (R) core (TM) *i*7-4700MQ CPU and 8GB of RAM. We use 10-fold cross-validation to estimate the performance of our proposed classifier. In a 10-fold cross validation, data are split into 10 equal folds of same sizes prior to carrying out 10 iterations of training and validation.

Table 1 Feature Selection using Filter Method

Filter method	Feature selected
Gain Ratio	12,26,4,25,39,6,30,38,5,29,3,37,34,33
Chi- Squared	5,3,6,4,29,30,33,34,35,12,23,38,25,39
ReliefF	3,29,4,32,38,33,39,12,36,23,26,34,40,31

From Table 1, it is seen that individual filter method has ranked the feature of the dataset according to its strength in determining the class. We have attached a weight to each ranking position, therefore, we determine the strongest features across the three filter methods, and cumulatively sum up the weights. Table 2 presents the output of our triple-filter method, that is, the fourteen most important features. These fourteen features have been used as the input for training the decision tree classifier, *J*48 in Weka.

Table 2 Triple-Filter Feature Selection Method

Filter method	Feature selected
Triple-filter	3,4,29,33,34,39,12,5,30,38,26,25,23,6

6. Performance Measure

During the evaluation of a classifier, different metrics such as classification accuracy, detection rate and false alarm rate can be used. These metrics are determinant on the measure of the true positive (TP), false positive (FP), true negative (TN) and false negative (FN). TP are the instances where attack packets are correctly classified, while situations of FP occur when certain amount of normal packets are misclassified as attack (false alarm). TN, on the other hand, are situations where normal packets are correctly classified, whereas, FN are instances where packets are classified as normal, when indeed they are attacks. Recently developed IDS for detecting attacks in WSN requires a relatively high detection rate with low false alarm. As discussed, in this work, we consider the classification

accuracy, detection rate and false alarm rate of our triple-filter method. We compare these metrics with the performance of the full dataset containing all the features and each of the filter methods using *J*48 classifier. The metrics used for comparison are defined as follows.

1. Classification accuracy: This is defined as the ratio of the data defined correctly to that of the entire dataset in percentage. The accuracy of a proposed technique can be derived using the formula:

$$CA = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \tag{7}$$

2. Detection rate: Detection rates is usually based on the confusion matrix and can be determined by using the formula

$$DR = \frac{TP}{TP+FN} \times 100\%. \tag{8}$$

3. False alarm rate: This is the amount of normal data that are misclassified as attack during detection. The false alarm rate can be determined by using the formula:

$$FAR = \frac{FP}{FP+TN} \times 100\% \tag{9}$$

Table 3 presents the performance measure of our proposed IDS defence solution with respect to the classification accuracy, detection rate, and false alarm rate.

7. DISCUSSION

Intrusion detection in WSN during an attack can further increase the complexity and resource consumption of the SNs. Thus, filter methods for feature selection when compared to wrapper methods are fast and easy to interpret. However, previous research has shown that it cannot determine features that are strong as a group but weak individually [39]. We have chosen to deploy our proposed IDS on the cluster heads because we assume that the cluster heads have better battery life with a higher software and hardware capability as compared to the other nodes. Fig. 3 shows the classification accuracy across different filter feature selection methods and our triple-filter method.

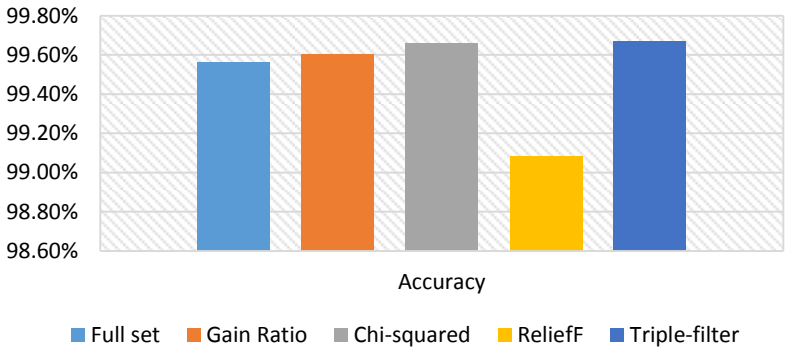

Fig. 3 Classification accuracy for different filter methods

As shown in Fig. 3 and Table 3, our proposed method exhibits the best accuracy performance. It presents a slight improvement of 0.01 % as compared to chi-squared filter method which gives the second best accuracy performance. In Fig, 4, the detection rate across the different filter methods and our proposed triple-filter is presented. The result shows that our proposed filter method with 14 selected features offers the best detection rate in comparison with the other filter methods. As shown in Table 3 and Fig. 4, the triple-filter method offer a slight increase in detection rate of 0.02% when compared with the next best filter feature selection method.

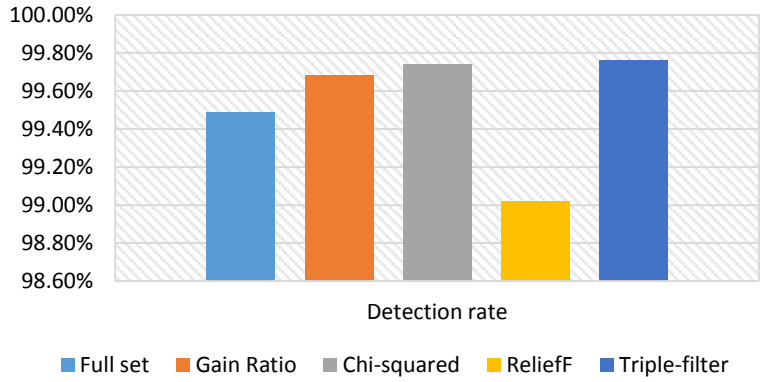

Fig. 4 Detection rate for different filter methods

For the false alarm rate, ReliefF presents the worst result of 0.87% while the full dataset (with the entire features) showcase the best performance, 0.38%. Our proposed method presents a false alarm rate of 0.42% as shown in Fig 5. Although, our proposed triple-filter method do not offer the best false alarm rate, it is still suitable in real-time WSNs because it offer good classification accuracy and detection rate at a reduced complexity. Note that a lightweight IDS is an important requirement in order to prolong the lifetime and functionality of sensor networks.

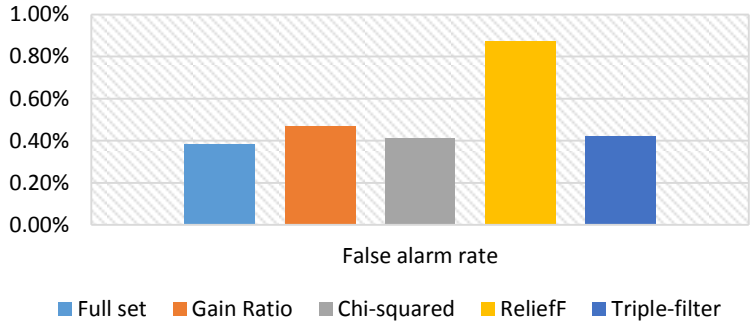

Fig. 5 False alarm rate for different filter methods

Table 3 Performance comparison of the triple-filter with full dataset, gain ratio, chi-square and ReliefF

Filter methods	No of features	Accuracy	Detection rate	False alarm rate
Full set	41	99.56%	99.49%	0.38%
Gain Ratio	14	99.60%	99.68%	0.47%
Chi-squared	14	99.66%	99.74%	0.41%
ReliefF	14	99.08%	99.02%	0.87%
Triple-filter	14	99.67%	99.76%	0.42%

Finally, we compared the triple-filter method with a similar work in [10]. Table 4 presents the performance comparison of the proposed triple-filter method with the work in [10].

Table 4 Performance comparison of the triple-filter with the work in [10]

Filter method	Classifier	No of features	Accuracy	Detection rate	False alarm rate
SVM-RFE [32]	BPN	24	99.75%	95.13%	2.06%
Triple-filter	J48	14	99.67%	99.67%	0.42%

As presented in Table 4, the triple-filter feature selection, with 14 features, present an improvement in the detection rate and the false alarm rate as compared with the work in [10] using the NLS-KDD dataset. This shows the efficiency of our proposed triple-filter feature selection method in improving the detection rate of the decision tree classifier with minimal false alarm while conserving the limited resources of the sensor network.

8. Conclusion and Future Work

In this paper, we have proposed the combination of three filter feature selection methods, gain ratio, chi-squared and ReliefF, called triple-filter, to pre-process dataset prior to attack classification. The proposed feature selection method is deployed in a heterogeneous cluster-based WSN, where the IDS is implemented on the cluster head nodes. The proposed IDS reduce the complexity of the system by selecting important features in the dataset, thus reducing the features from 41 to 14 before classification, using a decision tree algorithm, *J*48. Experimental results obtained show an improved performance with reduced feature set from 41 to 14. Also, our proposed triple-filter feature selection method performed better than individual filter methods using *J*48 classifier. In the future, we seek to extend our work to study the effect of our solution on homogeneous WSNs and also to evaluate our proposed triple-filter feature selection on other classification algorithms.

References

[1] I. Almomani, B. Al-Kasasbeh, M. AL-Akhras, "WSN-DS: A Dataset for Intrusion Detection Systems in Wireless Sensor Networks*", Journal of Sensors*, pp. 1–16, 2016.

[2] C. O'Reilly, A. Gluhak, M. A. Imran, S. Rajasegarar, "Anomaly detection in wireless sensor networks in a non-stationary environment", *IEEE Communications Surveys & Tutorials*, vol. 16, pp. 1413–1432, 2014.

[3] O.O. Ogundile, A. S. Alfa (2017), "A Survey on an Energy-Efficient and Energy-Balanced Routing Protocol for Wireless Sensor Networks", *Sensor*, vol. 17, 1084, 1–51, 2017.
[4] O. Osanaiye, A. Alfa, "Denial of Service Defence for Resource Availability in Wireless Sensor Networks", *IEEE Access*, vol. 6, pp. 6975–7004, 2018.
[5] H.M. Salmon, et al, "Intrusion detection system for wireless sensor networks using danger theory immune-inspired techniques", *International journal of wireless information networks*, vol. 20, pp. 39–66, 2013.
[6] V. F. Taylor, D. T. Fokum, "Mitigating black hole attacks in wireless sensor networks using node-resident expert systems", In Proceedings of the IEEE Wireless Telecommunications Symposium (WTS), pp. 1–7, 2014.
[7] S. Athmani, D.E. Boubiche, A. Bilami, "Hierarchical energy efficient intrusion detection system for black hole attacks in WSNs", In Proceedings of the IEEE World Congress Computer and Information Technology (WCCIT), pp. 1–5, 2013.
[8] O. Osanaiye, R. Choo, M. Dlodlo, "Distributed Denial of Service (DDoS) Resilience in Cloud: Review and Conceptual Cloud DDoS Mitigation Framework", *Journal of Network and Computer Applications*, vol. 69, pp. 1447–1465, 2016.
[9] M. Tavallaee, E. Bagheri, W. Lu, A. Ghorbani, "A detailed analysis of the KDD CUP 99 dataset", In Proceedings of the Second IEEE Symposium on Computational Intelligence for Security and Defence Applications CISDA, pp. 1–6.
[10] Wang S.-S., Yan K.-Q., Wang S.-C., Liu C.-W. (2011) An integrated intrusion detection system for cluster-based wireless sensor networks. *Expert Systems with Applications*, 38, 15234–15243.
[11] O. Osanaiye, H. Cai, K.K.R. Choo, A. Dehghantanha, Z. Xu and M. Dlodlo, "Ensemble-based multi-filter feature selection method for DDoS detection in cloud computing" *EURASIP Journal on Wireless Communications and Networking*, vol. 130, pp. 1–10, 2016.
[12] X. Deng, "An intrusion detection system for cluster based wireless sensor networks", In Proceedings of the 16th IEEE International Symposium on Wireless Personal Multimedia Communications (WPMC), 2013, pp. 1–5.
[13] K. Q. Yan, S.C. Wang, S.S. Wang, C.W. Liu, "Hybrid Intrusion Detection System for enhancing the security of a cluster-based Wireless Sensor Network", In Proceedings of the 3rd IEEE International Conference on Computer Science and Information Technology (ICCSIT), vol. 1, 2010, pp. 114–118.
[14] K. Medhat, R.A. Ramadan, I. Talkhan, "Distributed Intrusion Detection System for Wireless Sensor Networks", In Proceedings of the 9th IEEE International Conference on Next Generation Mobile Applications, Services and Technologies, 2015, pp. 234–239.
[15] M. Tiwar, K.V. Arya, R. Choudhari, K. S. Choudhary, "Designing intrusion detection to detect black hole and selective forwarding attack in WSN based on local information", In Proceedings of the 4th IEEE International Conference on Computer Sciences and Convergence Information Technology ICCIT'09, 2009, pp. 824–828.
[16] A.P.R. Da Silva, et al, "Decentralized intrusion detection in wireless sensor networks" In Proceedings of the 1st ACM international workshop on Quality of service & security in wireless and mobile networks, 2005, pp. 16–23.
[17] Jong K., Marchiori E., Sebag M., van der Vaart A. (2004) Feature Selection in Proteomic Patten Data with Support Vector Machines. *Symposium on Computational Intelligence in Bioinformatics and Computational Biology*, 41–48.
[18] J. Yick, B. Mukherjee, D. Ghosal, "Wireless sensor network survey", *Computer networks*, vol. 52, pp. 2292–2330, 2008.
[19] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, E. Cayirci, "A survey on sensor networks", *IEEE Communications magazine*, vol. 40, pp. 102–114, 2002
[20] J. Heidemann, et al, "Research challenges and applications for underwater sensor networking" In Proceedings of the IEEE Wireless Communications and Networking Conference, WCNC, 2006, pp. 228–235.
[21] I. F. Akyildiz, T. Melodia, K. R. Chowdhury, "A survey on wireless multimedia sensor networks", *Computer networks*, vol. 51, pp. 921–960, 2007.
[22] A. Abduvaliyev, A.-S. K. Pathan, J. Zhou, R. Roman, W.-C. Wong, "On the vital areas of intrusion detection systems in wireless sensor networks", *IEEE Communications Surveys & Tutorials*, vol. 15, pp. 1223–1237, 2013.
[23] Y. Yu, K. Li, W. Zhou, P. Li, "Trust mechanisms in wireless sensor networks: Attack analysis and countermeasures", *Journal of Network and computer Applications*, vol. 35, pp. 867–880, 2012.

[24] C.C. Su, K.M. Chang, Y.H. Kuo, M.F. Horng, "The new intrusion prevention and detection approaches for clustering-based sensor networks", In Proceedings of the IEEE Wireless Communications and Networking Conference, vol. 4, 2015, pp. 1927–1932.
[25] M. H. Anisi, A. H. Abdullah, S. A. Razak, "Energy-efficient and reliable data delivery in wireless sensor networks", *Wireless Networks*, vol. 19, pp. 495–505.
[26] P. Kuila, P.K. Jana, "Energy Efficient Load- Balanced Clustering Algorithm for Wireless Sensor Networks", *Procedia Technology*, vol. 6, pp. 771–777, 2012.
[27] P. Kuila, S. K. Gupta, P.K. Jana, "A novel evolutionary approach for load balanced clustering problem for wireless sensor networks", *Swarm and Evolutionary Computation*, vol. 12, pp. 48–56, 2013.
[28] P. Kuila, P.K. Jana, "Approximation schemes for load balanced clustering in wireless sensor networks", *Journal of Supercomputing*, vol. 68, pp. 87–105, 2014.
[29] R. Xie, X. Jia, "Transmission-Efficient Clustering Method for Wireless Sensor Networks Using Compressive Sensing", *IEEE Trans. Parallel Distrib. Syst*., vol. 25, pp. 806–815, 2014.
[30] O. A. Osanaiye, DDoS defence for service availability in cloud computing. *Doctoral dissertation, University of Cape Town*, 2016.
[31] V. Bolon-Canedo, N. Sanchez-Marono, A. Alonso-Betanzos, "A review of feature selection methods on synthetic data", *Knowledge and information systems*, vol. 34, no. 3, pp. 483–519, 2013.
[32] C. J. Mantas, J. Abellan, "Credal-C4. 5 Decision tree based on imprecise probabilities to classify noisy data", *Expert Systems with Applications*, vol. 41, pp. 4625–4637, 2014.
[33] H.F. Eid, A.E. Hassanien, T.H. Kim, S. Banerjee, "Linear correlation-based feature selection for network intrusion detection model", *In Advances in Security of Information and Communication Networks*, pp. 240–248, 2013.
[34] M.B. Yassein, Y. Khamayseh, M. AbuJazoh, "Feature Selection for Black Hole Attacks", *Journal of Universal Computer Science*, vol. 22, no. 4, pp. 521–536, 2016.
[35] J. Gehrke, V. Ganti, R. Ramakrishnan, W.Y. Loh, "BOAT-optimistic decision tree construction" In ACM SIGMOD Record, vol. 28, pp. 169–180, 1999.
[36] N. Sanchez-Marono, A. Alonso-Betanzos, M. Tombilla-Sanroman, "Filter methods for feature selection - a comparative study", *Intelligent Data Engineering and Automated Learning-IDEAL*, pp. 178-187, 2007.
[37] N. Sengupta, J. Sen, J. Sil, M. Saha, "Designing of on line intrusion detection system using rough set theory and Q-learning algorithm", *Neurocomputing*, vol. 111, pp. 161-168, 2013.
[38] http://www.cs.waikato.ac.nz/ml/weka/, [Online] access 2nd August 2017.
[39] O. Osanaiye, R. Choo, M. Dlodlo, "Analysing feature selection and classification techniques for DDoS detection in cloud", In Proceedings of the Southern Africa Telecommunication, pp. 198-203, 2016.