International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-8 Issue-11, September 2019

Feature Reduction using Lasso Hybrid Algorithm in Wireless Intrusion Detection System

D.Sudaroli Vijayakumar, Sannasi Ganapathy

Abstract: ***To maintain the integrity and protection of networks, intrusion detection systems play a vital role. Growth of wireless networks turned the globe to perform all pecuniary tasks online resulting a lot of security breaches in the network. One of the common breaches happening in network is the intruders who eventually tries to bypass the adopted security framework. Every day new intrusions arises and new solutions as well, however the research in making the intrusion detection system intelligent holds energetic. Today most of the systems are becoming intelligent by adopting machine learning and artificial intelligence algorithms. Success of building an efficient machine learning model to make intelligent intrusion detection system is relied on the effective features considered for classification and prediction. Thus, feature reduction is an integral part for discarding irrelevant and redundant features to produce a computationally decisive system that can identify defects with high accuracy. This implementation is an attempt to identify the smaller feature set possible for the well adopted wireless intrusion detection dataset AWID. Here, we proposed a LASSO based implementation to produce a smaller decisive set of features. Incorporation of Lasso on feature reduction not only provides a smaller set of features, but also allow to adopt prediction algorithms inside Lasso resulting lesser number of false alarms as well.***

Keywords: WLAN 802.11, Intrusion Detection System (IDS), Machine Learning, Enhanced Feature Selection (EFS), Random Forest, Lasso, Wireless IDS (WIDS).

I. INTRODUCTION

Over the recent years, the usage of internet is dramatically increased, and it has led to a momentous inflation in the amount of data generated by the wireless networks. According to Worldwide telecommunication consortium [1], this amount of data is going to be increased 1000 folds, the major contributor of holding 66% comes from the wireless networks. Higher the number of networked devices, higher the types of attackers and attacks. Although wireless networks provided convincing results in terms of mobility, confidentiality and sharing, security rift prevails. Intrusion Detection System is a mechanism followed both in wired and wireless networks to address this security rift. Despite all the efforts, the intrusion detection system fails to behave intelligently because of the amount of data it has to analyze for discriminating the usual from unusual. With a huge amount of traffic and features Wireless Intrusion Detection system must compromise on integrity.

To obtain better prediction results, minimal as well as best features should be considered. Selection of ideal features is the greatest difficulty while building prediction model for WIDS. The success of classification accuracy is relied on the lower number of features selected for classification. There are lots of extensive research in order to make the number of features for classification and prediction as low as possible. Feature selection or reduction tries to create a target classification variable by creating a subset of features retaining the original feature [2].

A wireless Local Area Network (WLAN), type of network does not hold physical connection. Instead it uses high frequency radio waves to connect network devices and its wireless range is measured in hundreds of meters. To safeguard WLAN against security issues, it is essential to keep track of all the access points, actions to close if any unauthorized access points found, to know what users and unencrypted data is exchanged. To account the above-mentioned information, WIDS play a major role. In their simplest form, any attack that disrupts the normal functioning of surveillance framework can be termed as intrusion. Detecting intrusion is the course of scanning and scrutinizing the various happenings of the network to detect deformity [3]. WIDS are designed to capture and inspect the list of Access points as well as signal strength and transmission speed. Intrusions are categorized depending on two strategies [4]: misuse detection and anomaly detection. Misuse detection strategy looks for network attack sequences that match a predefined pattern. This is achieved through the maintenance of periodically updated database that stores the signatures of known attacks. Unless trained, misuse strategy cannot detect new attacks. Any new attacks can be effectively identified as it tries to find any deviations from the normal behavior. The so-called normal behavior is defined by the network administrator. The anomaly detection is flexible enough to deal with new attack.

General classification that affects the network are grouped among the below classes:

- Probe Attacks: The target for the attackers in this type of attacks is the device itself and all information about the device is affected.
- Denial-Of-Service (DoS): Attackers force the target to stop the services.
- Remote-to-Local(R2L): Attackers tries to access the target machine without having valid account.
- User-to-Root(U2R): With the native privilege attackers tries gain super user privileges.

Manuscript published on 30 September 2019.
***Correspondence Author(s)
D.Sudaroli Vijayakumar*, Computer Science, PES University, Bangalore, India. Email: oli.sudar@gmail.com
Sannasi Ganapathy, Computer Science, VIT University, Chennai, India. Email: sganapathy@vit.ac.in

One of the common steps when dealing any problem statement using machine learning approach is feature selection. Leaning algorithm gives better accuracy if the algorithm can visualize the training set easily that is supported with fewer features. The objective of deriving valuable insights from data is successful if the statistical measures can give accurate results. Statistical metrics demands reduction with volume space retaining good sparsity. Hence the need for selecting a good subset of the original features is unavoidable for faster and efficient detection. Thus prior to the application of learning algorithm, an essential data processing step is feature selection.

Feature reduction can be performed by adopting wrapper-based feature reduction, that requires the application of a single algorithm that can perform learning as well providing smaller set of features. A filter-based feature reduction aims to select subset considering the general characteristics of data. Although computationally expensive, Wrapper based feature reduction provides better accuracy because of its ability to add or remove features from the subset seeing the inferences. Any feature selection methodology adopted should provide these benefits of reduction in overfitting, improvement in accuracy and reduction in training time. Each of the above-mentioned advantage has proximity with each other like the model will have low accuracy if it is overfitting and so on. Any technique that helps in avoiding overfitting and increasing model interpretability would be a great choice for effective feature reduction. Here the process of selecting features is via an embedded selection strategy that involves LASSO that addresses the problem of overfitting, accuracy and training time yielding a better feature in terms of number.

The objective of building a statistical model is to explain the data as close as possible to reality. Any model comprises of two types of variables dependent and independent. Good model is the result of choosing the right independent variables that influence the dependent variable. The selection of features gains greater importance particularly in high dimensional dataset. To increase the prediction accuracy and interpretability of such a large dataset can be addressed with the help of LASSO. LASSO standing for Least Absolute Shrinkage and Selection Operator is a regularization and good strategy in selecting best features for statistical models. Models can provide better results if the derivation of target output is limited with lesser features.

Lasso stands unique with its objective function for minimization and it includes regularizer.

The lasso estimate solves the minimization of the least-squares penalty with α || w ||1 added,

$$\hat{\beta}^{lasso} = arg\min_{\beta} \sum_{i=}^{n} (y - (\beta + \beta^T x)) + \alpha||\beta||$$

where α is a constant and ||w||1 is the l_1-norm of the parameter vector. Lasso with its iterative property tries to identify the best contributing feature. The tuning parameter λ determines the shrinking variable. Larger value of λ means the coefficients are forced to be exactly equal to zero so there are a greater number of coefficients are shrinked to Zero. Cross-validation would be the optimal way of identifying the penalty factor. The algorithm used for finding the minima of the function is Coordinate descent because computing gradients for large amounts of features is very resource-heavy task. With the reduction in volume space, Lasso regression can effectively the dearth coefficients using alpha guideline. In this paper, we are using this LASSO that will serve as an evolutionary model to continuously reduce features after the application of learning algorithm.

In this paper, the machine learning feature reduction algorithms will be evaluated on the AWID, well known wireless Intrusion Detection Dataset. The feature reduction is dealt in three perspectives (i) removing the single, missing and co-linearity values by applying straight forward search and replace algorithm (ii)Application of standard machine learning feature reduction algorithms like EFS and Random forest (iii) Application of Lasso with hybrid algorithm called HERLA. The HERLA algorithm results serves as a baseline to perform comparative analysis among the available reduction methods and to show that this algorithm holds good for AWID feature reduction. Rest of the paper is organized as follows: Section 2 give overview on the related works of Wireless Intrusion Detection System and machine learning feature reduction methods. Section 3 describes about system framework. Section 4 describes the proposed method. Method is validated with result analysis is portrayed in Section 5. Scope of further improvement and conclusions are drawn in section 6.

II. RELATED WORKS

The various effective Intrusion Detection Systems [35-39] and intelligent data mining systems [40-45] are perceived and discussed by eminent educators in this direction. The conception of intrusion detection started with the work by Anderson [5] that described misuse detection is possible through the audit trails. The problem of providing a better machine learning model to predict the future attacks appropriately is dealt by researchers by following supervised or unsupervised methods. Aminanto [6] briefly gone through the opportunities of using the varied learning methodologies in intrusion dataset. He claims supervised method suits well for labeled data and unsupervised holds good for unlabeled data. Generally, the benchmark wireless intrusion detection dataset AWID is unlabeled, so the suitability of using unsupervised learning methodologies holds priority. Most Commonly used unsupervised learning methods such as *k*-means clustering [7], Principal Component Analysis (PCA) [8] and Independent Component Analysis (ICA) [9] cannot handle wireless intrusion dataset effectively as the data is not well distributed [10]. This problem was addressed effectively by Aminanto [6] with SAE, a deep learning method that transforms the original features to meaningful representation thus improving the results produced by unsupervised learning methods. The amalgamation of deep learning with k-means clustering have achieved 92% detection rate for impersonation attack. Kayacik [13] explains the relevance to select features in Intrusion detection system with KDD 99 Dataset. Manekar and Waghmare [16] achieved an optimal set of features with unique approach of combining the Practical Swarm Organization (PSO) and support vector machine (SVM) in which SVM performed the classification task and PSO did feature optimization. Guyon [19] attempted to choose features using SVM. Zaman and Karray [14] proposed “Enhanced Support Vector decisive function” that selects features based on priority and interrelationship among them.

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

Usha [21] used the AWID dataset and applied SVM methodology to detect the anomalies in WIDS.

The problem of feature reduction is approached using the Bayesian concepts by several authors. Automatic feature selection is obtained using the Correlation based feature selection in [22]. In [23] a different perspective of defining a composite mechanism using Bayesian network and Classification and regression tree performances (CART) to identify best features is identified. An ensembled version of Bayesian and CART is finally been presented as a hybrid architecture. [25] used the Bayesian rule of conditional probability to denote the base rate fallacy for intrusion detection. In [26] a model is built using Bayesian technique to retrieve features.

Another completely different dimension of dealing with intrusion dataset is observed in [28] wherein a commendable effective pattern representation and dimensionality reduction is achieved in CICIDS2017 dataset. AE and PCA are used to obtain dimensionally reduced features and this efficiency of dimensionally reduced features is verified using popular classification algorithms. Generally, the light weighted features are eliminated while applying feature selection methods that contribute for intrusion. Bio-inspired deep learning algorithm like Auto Encoder can make a huge difference when dealing with intrusion detection due to the dynamic network traffic. [29] presents the methodology of using auto encoder algorithm to perform not only feature extraction and dimensionality reduction but also classification and clustering.

A remarkable literature in this area comes in the form of [30], in which an accretion clustering is adopted. The concept of incremental clustering holds potentiality in dealing influential data. Incremental clustering is not a new concept in all the traditional clustering algorithms, however in this, a fast outlier detection algorithm by combining the K-means algorithm and local density score is remarkable. One of the major issues faced while dealing with network intrusion dataset is the isolation of data sets specific to scenarios. In this paper [31] the authors have identified the different properties to produce a well labelled data set with a clear-cut explanation on the distinctiveness of the data on specific scenarios. The overall 15 properties are grouped into 5 categories that can serve as benchmark to choose and work on the features specific to problem formulation.

One technology that is taking a paradigm shift is artificial intelligence. In this work [32]a completely different dimension on handling the intrusion dataset is presented. Convolutional Neural Network (CNN) often producing convincing results in imaging and voice analysis. The idea of using CNN directly on to the intrusion dataset might not produce better result. So “making data fit model” is created using data visualization onto this dataset. This is the only work in literature showing convincing results in all the four categories of attacks.

Qin, Y., [33] has proposed the first two-dimensional data cleaning method in which a remarkable reduction of features was made possible. Out of 154 attributes, 18 useful attributes are selected and applied SVM to denote abnormalities and have achieved the detection accuracy for the various classification attacks ranging from 85% to 99%. Ensemble based learning is used in [34] to identify the necessary features for efficient intrusion detection system implementation. Various ensemble-based methods are studied with the performance metrics like accuracy, precision, recall and f-measure and their respective results are compared to identify the best ensemble method suitable for feature reduction in AWID dataset.

The standard benchmark dataset called KDD’99 Cup dataset has been used for evaluating the intrusion detection system in the various intrusion detection systems [35-39] and the intelligent agents are also used by various researchers in their prediction and detection systems [40-45]. These intelligent systems have been achieved better detection accuracy with less false alarm rate. Even though, these works are not providing the users expected detection accuracy over their systems. With this literature survey, some of the points that remain as an issue while dealing with intrusion detection dataset are:

i. Inability to predict all the four categories of attacks effectively.
ii. Reduced features will produce better prediction for all four categories of attacks.
iii. Relationship between dimensionality reduction and prediction.

To build a highly decisive and human like detection system, selecting effective attributes is a crucial step. Since the feature selection has direct impact on the classifier, it is important to identify an effective methodology that can continuously reduce the features even during the learning phase.

This work considered the AWID dataset is initially tried to reduce the number of features using the standard EFS and Random forest. Since the number of features, we got as a resultant of these algorithms is not satisfactory, a hybrid algorithm combining the idea of EFS and random forest was implemented, and the reduction was tested. This algorithm gave satisfactory result however the idea is to build a system that can continuously reduce the features in classification and prediction as well. One model that can perform this process is LASSO. The LASSO is implemented along with the Navies Bayes Classifier and regression methodology.

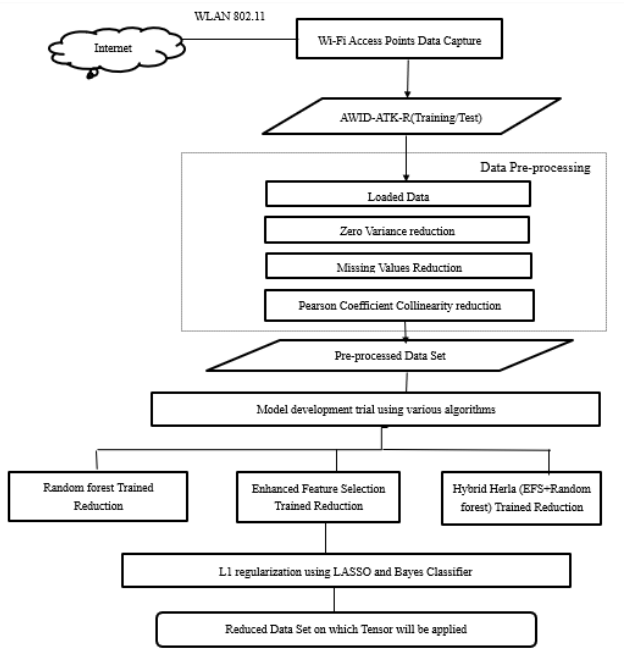
III. SYSTEM FRAMEWORK

This section introduces the system model for our proposed feature reduction method, which can be further used effectively to detect intrusion detection as well as in reducing the number of false alarms. This feature reduction method is based on LASSO and Bayes Classifier. Fig 1 illustrates the system framework for evolving feature reduction. This system has two main components: data preprocessing and feature reduction.

This paper uses AWID dataset for intrusion detection. This dataset consists of 155 features in which many of the attributes prove misleading. Reliable data must be extracted to perform better prediction. Thus, preprocessing is carried out carefully to produce a reliable set of data. The process of preprocessing involves single value elimination, missing values and the existence of linear relationship between attributes.

Fig:1 System Framework for Evolving Feature Reduction

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)



The preprocessed data can be further reduced for its features if suitable feature reduction algorithm is used. With this idea, feature reduction algorithms in the form of random forest, enhanced feature selection was applied, and the total number of features was measured by carefully observing the Pearson correlation coefficient value. Since these models didn't found satisfactory results, an algorithm called HERLA is adopted that combined some of the features of random forest and enhanced feature reduction algorithms. Deployment of HERLA gave satisfactory reduction that was used as a base and an evolutionary feature reduction approach called LASSO was used to continuously learn and reduce features depending on the learning algorithm used. The adopted technology can work as a better feature reduction model as it is going to continuously reduces the features along with the classification and prediction.

IV. PROPOSED METHOD

The core idea of obtaining better prediction accuracy and reduction in false alarms is only possible with good subset and as shown in Fig.1, the data must be processed. The following section focuses on this aspect of the study. The process of feature reduction is carried out through the steps including Feature reduction includes cleaning, correlation analysis and preprocessing and Lasso feature reduction using Bayes Classifier. The basic principles of each step are described below.

4.1 Feature Extraction and Selection

The AWID-ATK-R-Trn dataset was used to train the machine learning techniques and the AWID-ATK-R-Tst dataset was used to evaluate the performance of the seven classifier models. In this paper, we use the AWID-CLS-R-Trn dataset for training and the AWID-CLS-R-Tst dataset for testing. In this approach, we have achieved the subset of 68 features by reducing the AWID dataset of 155 features. Upon initial glance of the data, one of the first things noticed is the number of features that have a single unique value in all its rows. This is not very useful as it provides zero variance in the data and can be safely dropped. The data upon inspection, was found to have 52 features with a single unique value. It is also important to remove single unique values because a tree-based model like Random Forest can never make a split on a feature with only one value in the entire row. These types of values can be removed by using a straightforward search-and-replace algorithm without much fear of data loss.

The data was found to have 31 features with greater than 95% missing values. To handle this missing data, we can either drop columns, fill the missing values, or drop the whole feature.

Most machine learning algorithms assume that features fed to them are not multicollinearity data. Thus, we must drop features with high correlations. Highly correlated features can be identified by using the Pearson correlation coefficient. Pearson's correlation coefficient is the covariance of the two variables divided by the product of their standard deviations:

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2(y_i - \bar{y})^2}$$

where x, y are the pair of variables from comparing features. The dataset contains 14 features with correlation threshold greater than 0.95. Every node in the decision trees is a condition on a single feature, designed to split the dataset into two so that similar response values end up in the same set. The measure based on which the optimal condition is chosen is called node impurity. By deleting trees below a node score, we can create a subset of the most important features.

Calculating impurity can be done using the Gini impurity formula:

To compute Gini impurity for a set of items with J classes, suppose i ∈ { 1 , 2 , . . . , J }, and let p(i) be the fraction of items labeled with class i in the set:

$$Gini(E) = 1 - \sum_{cj=1} p2j$$

Input:
F= full set of 155 features
ac= classifier accuracy

//Build decision Model based on $Gini(E) = 1 - \sum_{cj=1} p2j$

//LASSO BASED IMPLEMENTATION
Initialize: S={F}
For each feature {f} from
(1) Iterating all columns and selecting only features with single value
(2) Iterating all columns and selecting only features that exceeds the specified % of missing Values
(3) Iterating over all columns and selecting only two features
(4) Iterate on all columns and calculate Gini Impurity
(5) Build decision tree based on the impurity value and split branches according to desired importance parameter 0.05
(6) Using Lasso displayed Gini importance through Class Variable
F=S> threshold // Set of reduced Features
End

When training the tree, it has simultaneously been computed how much each feature decreases the weighted impurity in a tree. 13 Features have an importance rating of lower than 0.05 and can be dropped.

Lasso is a meta-transformer that can be used along with any estimator that has a coef_ or feature_importances attribute after fitting. The features are considered unimportant and removed, if the corresponding coef_ or feature_importances values are below the provided threshold parameter. This is selected it because it has built-in variable selection by taking the coefficient of features found to be close to 0, and decision trees which displays Gini importance through a class variable. Upon using a threshold value of 0.2x of the mean, it was possible to select features based on the order of each one's Gini score.

V. EXPERIMENTAL RESULT AND ANALYSIS

5.1 Experimental Setup

We have used the python NumPy and panda package to compute the minimal subset of features and to extract the relevant features. We have used the existing implementation of a meta transformer available in the scikit-learn python library.

5.2 Dataset

The dataset we used in our experiments was AWID Aegean Wi-Fi Intrusion Dataset AWID [6] that mimics a common WLAN scenario. The AWID dataset comprises of five days network traffic and it comes in two forms mentioned as

i) "\CLS (Attack Class):

Each record in this class is labelled as attack classes and the attack classes include flooding, impersonation and injection and the fourth class denoted the normal class. These target classes are mapped onto the integer valued classes: 1 for normal instances, 2 for impersonation, 3 for flooding and 4 for injection attacks.

ii) "\ATK" (Attack Specific):

The attack-specific version labels the same traffic records with one of 17 different Wi-Fi attacks or normal.

We have used the reduced version of the AWID dataset. Table 1 and 2 shows the distribution of records

Table 1 shows the types of AWID dataset

Class	Training	Test
AWT-CLS-R-TRN	1,633,190	530,785
AWT-CLS-R-TST	1,63,190	53,078
AWT-ATK-R-TRN	1,62,385	
AWT-ATK-R-TST		44,858

Table 2 shows the number of Attack Classes in AWID

Class	Training	Test
Impersonation	48,552	20,079
Flooding	48,484	8,097
Injection	65379	16,682
Total	1,62,385	44,858

Each record in the dataset possess 155 features that represent the values for various header fields present in each frame. Features are essential to identify the behavior of the network traffic, though all features are relevant in identifying the specific type of attacks few features are not usable by any means and thus can be termed as noise. These noise-like features must be removed for better classification accuracy.

The data cleaning kicked off with the standard implementation of the Enhanced Feature Selection Algorithm through which the fraction of missing values above a specific threshold is identified and it showed 90% missing values for nearly 43 features. The sample features are depicted in the below table.

Table 3 shows some of the attributes showing above 90% missing values

Wlan.tkip.extiv	0.975266
Wlan_mgt.rsn.capabilities.preauth	0.942311
Wlan_mgt.rsn.akms.type	0.942311
Wlan_mgt.rsn.capabilities.peerkey	0.942311
Wlan_mgt.rsn.capabilities.mfpc	0.942311
Wlan_mgt.rsn.capabilities.mfpr	0.942311
Wlan_mgt.rsn.capabilities.gtksa_replay_counter	0.942311
Wlan_mgt.rsn.akms.count	0.942311
Wlan_mgt.rsn.capabilities.ptksa_replay_counter	0.942311
Wlan_mgt.rsn.version	0.942311
Wlan_mgt.rsn.gcs.type	0.942311
Wlan_mgt.rsn.pcs.count	0.942311
Wlan_mgt.rsn.capabilities.nopairwise	0.942311
Wlan.qos.bit4	0.9259513
Wlan.qos.txop_dur_req	0.9259513
Wlan_mgt.tim.bmapctl.offset	0.910203
Wlan_mgt.tim.bmapctl.multicast	0.910203
Wlan_mgt.tim.dtim_count	0.910203

Collinear features are features that are highly correlated with one another. In machine learning, these lead to decreased generalization performance on the test set due to high variance and less model interpretability. Thus, the enhanced feature selection algorithm is used again to extract those attributes possessing collinear features. 14 attributes were identified that is having a correlation of magnitude greater than 95%.

Table 4 shows attributes with correlation magnitude above 95%

	Corr_feature	Corr_value	Drop_feature
0	Frame.time_delta	1.0	Frame.time_delta_displayed
1	Frame.time_epoch	1.0	Frame.time_relative
2	Frame.len	1.0	Frame.cap_len
3	Radiotap.length	1.0	Radiotap.present.tsft
4	Radiotap.length	1.0	Radiotap.present.flags
5	Radiotap.present.tsft	1.0	Radiotap.present.channel
6	Radiotap.length	1.0	Radiotap.present.channel
7	Radiotap.present.tsft	1.0	Radiotap.present.channel
8	Radiotap.present.flags	1.0	Radiotap.present.dbm_antsignal
9	Radiotap.length	1.0	Radiotap.present.dbm_antsignal
10	Radiotap.present.tsft	1.0	Radiotap.present.dbm_antsignal
11	Radiotap.present.flags	1.0	Radiotap.present.dbm_antsignal
12	Radiotap.present.channel	1.0	Radiotap.present.antenna
13	Radiotap.length	1.0	Radiotap.present.antenna

After the cleaning process, the cleaned and memory-fixed dataset is loaded for further preprocessing in which again single unique elimination and collinear features removed totally 45 features with greater value of accuracy. As a part of preprocessing, Label Encoder is initialized for encoding the non-numeric data. After the application of Lasso and classifier 68 features are obtained

Table 5 shows sample features selected after Lasso and Classifier

```
Index(['frame.time_epoch', 'frame.time_delta', 'frame.len', 'radiotap.length',
       'radiotap.flags.cfp', 'radiotap.flags.preamble', 'radiotap.flags.wep',
       'radiotap.flags.frag', 'radiotap.flags.fcs', 'radiotap.flags.datapad',
       'radiotap.flags.badfcs', 'radiotap.flags.shortgi', 'radiotap.datarate',
       'radiotap.channel.freq', 'radiotap.channel.type.turbo',
       'radiotap.channel.type.cck', 'radiotap.channel.type.2ghz',
       'radiotap.channel.type.5ghz', 'radiotap.channel.type.passive',
       'radiotap.channel.type.dynamic', 'radiotap.channel.type.gfsk',
       'radiotap.channel.type.gsm', 'radiotap.channel.type.sturbo',
       'radiotap.channel.type.half', 'radiotap.channel.type.quarter',
       'radiotap.dbm_antsignal', 'radiotap.antenna',
       'radiotap.rxflags.badplcp', 'wlan.fc.type_subtype', 'wlan.fc.subtype',
       'wlan.fc.ds', 'wlan.fc.frag', 'wlan.fc.retry', 'wlan.fc.pwrmgt',
       'wlan.fc.moredata', 'wlan.fc.protected', 'wlan.duration', 'wlan.ra',
       'wlan.da', 'wlan.ta', 'wlan.sa', 'wlan.bssid', 'wlan.seq',
       'wlan.fcs_good', 'wlan_mgt.fixed.capabilities.ess',
       'wlan_mgt.fixed.capabilities.ibss',
       'wlan_mgt.fixed.capabilities.cfpoll.ap',
       'wlan_mgt.fixed.capabilities.privacy',
       'wlan_mgt.fixed.capabilities.preamble',
       'wlan_mgt.fixed.capabilities.pbcc',
       'wlan_mgt.fixed.capabilities.agility',
       'wlan_mgt.fixed.capabilities.spec_man',
       'wlan_mgt.fixed.capabilities.short_slot_time',
       'wlan_mgt.fixed.capabilities.apsd',
       'wlan_mgt.fixed.capabilities.radio_measurement',
       'wlan_mgt.fixed.capabilities.dsss_ofdm',
       'wlan_mgt.fixed.capabilities.del_blk_ack',
       'wlan_mgt.fixed.capabilities.imm_blk_ack', 'wlan_mgt.fixed.timestamp',
       'wlan_mgt.tagged.all', 'wlan_mgt.ssid', 'wlan_mgt.ds.current_channel',
       'wlan_mgt.tim.bmapctl.multicast', 'wlan_mgt.tim.bmapctl.offset',
```

5.1 Evaluation Metrics

The trivial algorithms helped us to develop the model with 128 features that we finalized in the data preparation phase. Then the implementation of Lasso helped us to derive 68 features, however the number of features is going to be decreased, since this is an evolutionary learning model. We attempted using the random forest classifier and Bayes Classifier. The Lasso with Bayes classifier provided us better results than that of the random forest. Though it is not decided as our final model as other learning algorithms like CNN can be used to achieve better accuracy. The performance of the selected model is analyzed using error matrix, F1 score and classification rate.

Error Matrix: Error Matrix are generally represented by means of classes in which every class represents the count values of the number of correct and incorrect predictions giving us an insight on the type of errors that are being made.

Classification Rate: It is defined on the ratio of the number of correctly predicted records and number of total number of records. Our classification rate as per the observation from confusion matrix is 94.12%

F-measure:

Since we have two measures (Precision and Recall) it helps to have a measurement that represents both. We calculate an F-measure which uses Harmonic Mean in place of Arithmetic Mean as it punishes the extreme values more. The F-Measure will always be nearer to the smaller value of Precision or Recall.

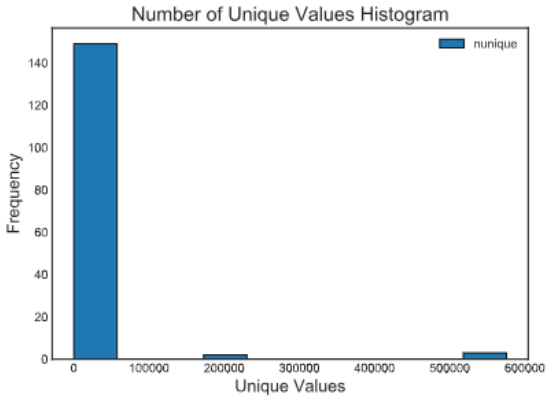
$$F-measure = \frac{2 \times recall * precision}{recall + precision}$$

Our model achieved 94.12% accuracy for identifying the traffic records with their classes.

5.3 Experimental Results

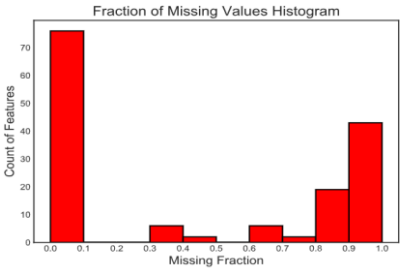
The results obtained after every step mentioned in the algorithm is presented in the form of graph.

Figure 1: Single unique Elimination after Trivial Algorithm



Single unique elimination is carried out using trivial algorithm by simply iterating over all columns to select all the features having only a single value. The above graph depicts the number of unique values keeping frequency and unique values. So, it is depicting the total number of features as 54.

Figure: 2 Features After Missing Fraction



The graph above shows the fraction (between 0 and 1) of missing values for the distribution of columns. The data was found to have 31 features with greater than 95% missing values. To handle this missing data, we can either drop columns, fill the missing values, or drop the whole feature.

Collinear features are features that are highly correlated with one another. In machine learning, these lead to decreased generalization performance on the test set due to high variance and less model interpretability. Collinearity reduction before applying the corresponding EFS and after applying EFS is depicted in figure 3 and 4.

Figure:3 All Correlations without applying any Collinearity

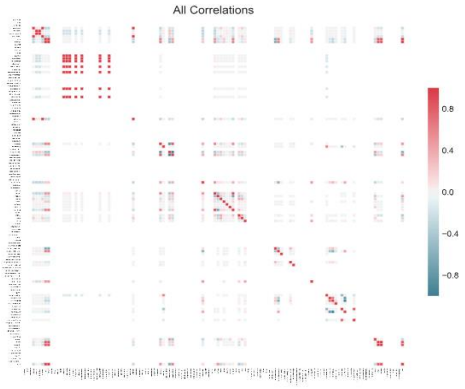
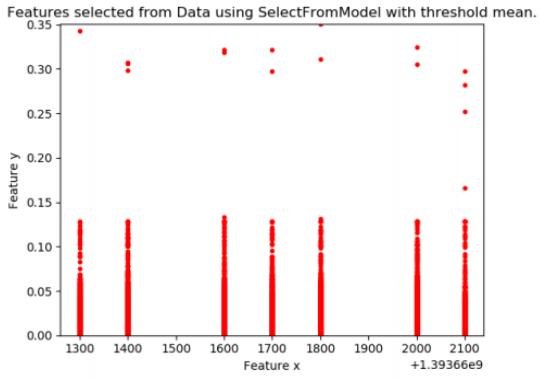


Figure 4: Collinearity after Pearson Correlation

Figure 5 shows the results after the application of random forest classifier and its efficiency is not satisfiable and an evolutionary approach like Lasso is required to continuously reduce the number of features. Figure 6 shows the final number of features with Lasso on Bayes Classifier.

Figure:5 Correlatin after Random Forest

Figure:6 Lasso with Bayes Classifier



The empirical results in Figure 6 clearly indicate that the total set of 155 features is reduced to 68 features with the Bayes classifier used Lasso.

VI. CONCLUSION AND FUTURE WORK

In this paper, we proposed Lasso Model for feature selection and made its comparison along with the random forest and Enhanced Feature Selection. Experimental results illustrate feature subset identified by Lasso has improved the number of features selected when compared to random forest, EFS and hybrid model of random forest and EFS. Since Lasso comes up with continuous feature reduction as we include the learning algorithm, future work will include this implementation along with a deep learning algorithm on top of Lasso to obtain a lesser subset of the ones which we obtained in this methodology. The deep learning algorithm can produce efficient results with an optimal feature, and this can also be tested replacing Lasso with elastic net.

REFERENCES

1. A. Osseiran, F. Boccardi, V. Braun, K. Kusume, P. Marsch, M. Maternia, O. Queseth, M. Schellmann, H. Schotten, H. Taoka, H. Tullberg, M. A. Uusitalo, B. Timus, and M. Fallgren, "Scenarios for 5G mobile and wireless communications: The vision of the metis project," IEEE Communications Magazine,52(5),26-35,2014.
2. R.kohavi,G.HJohn,Wrappers for feature subset selection, Artificial Intelligence,97(1-2),273-324,1997.
3. Base,R.Intrusion Detection,Macmillian Technical Publishing,2000.
4. Markou.M.Singh,S.,Novelty Detection:A review,Part I: Statistical Approaches,Signal Processing,8(12),pp.2481-2497),2003.
5. James P.Anderson. Computer Security Threat Monitoring and Surveillance,1980. http://csrc.nist.gov/publications/history/ande80.pdf
6. Aminanto, M.E., Kim, K.: Detecting impersonation attack in WiFi networks using deep learning approach. In: Choi, D., Guilley, S. (eds.) WISA 2016. LNCS, vol. 10144, pp. 136–147. Springer, Cham ,2017.
7. Jain, A.K.: Data clustering: 50 years beyond *k*-means. Pattern Recogn. Lett. **31**(8), 651–666 (2010)
8. Abdi, H., Williams, L.J.: Principal component analysis.Wiley Interdisc. Rev. Comput. Stat. **2**(4), 433–459 (2010)
9. Lee, T.W.: Independent Component Analysis, pp. 27–66. Springer, Heidelberg (1998).
10. Jiang, C., Zhang, H., Ren, Y., Han, Z., Chen, K.C., Hanzo, L.: Machine learning paradigms for next-generation wireless networks. IEEE Wirel. Commun. **24**(2),98–105 (2016)
11. Song, C., Liu, F., Huang, Y., Wang, L., Tan, T.: Auto-encoder based data clustering. In: Ruiz-Shulcloper, J., Sanniti di Baja, G. (eds.) CIARP 2013. LNCS,vol. 8258, pp. 117–124. Springer, Heidelberg (2013).
12. Aminanto, Harry Chandra Tanuwidjaja, Paul D Yoo,M.E., Kim, K.: Wi-Fi Intrusion Detection Using Weighted-Feature Selection for Neural Networks Classifier. In: Choi, D., Guilley, S. (eds.) IWBIS 2017.
13. H. G. Kayacik, A. N. Zincir-Heywood, and M. I. Heywood, "Selecting features for intrusion detection: A feature relevance analysis on KDD99intrusion detection datasets," *Proceedings of the Third Annual Conference on Privacy, Security and Trust*, Citeseer, 2005.
14. S. Zaman and F. Karray, "Lightweight IDS based on features selection and IDS classification scheme," *Proceeding of International Conference on Computational Science and Engineering, 2009*, vol. 3, pp. 365–370, IEEE, 2009.
15. P. Louvieris, N. Clewley, and X. Liu, "Effects-based feature identification for network intrusion detection," *Neurocomputing*, vol. 121, pp. 265–273, Elsevier, 2013.
16. V. Manekar and K. Waghmare, "Intrusion detection system using support vector machine (SVM) and particle swarm optimization (PSO)," *International Journal of Advanced Computer Research*, vol. 4, no. 3, p. 808, Accents, 2014.
17. H. Saxena and V. Richariya, "Intrusion detection in KDD99 dataset using SVM-PSO and feature reduction with information gain," *International Journal of Computer Applications*, vol. 98, no. 6, Foundation of Computer Science, 2014.
18. E. Schaffernicht and H.-M. Gross, "Weighted mutual information for feature selection," *International Conference on Artificial Neural Networks*, pp. 181–188, Springer, 2011.
19. I. Guyon, J. Weston, S. Barnhill, and V. Vapnik, "Gene selection for cancer classification using support vector machines," *Machine Learning*, vol. 46, no. 1-3, pp. 389–422, Springer, 2002.
20. Z. Wang, "The applications of deep learning on traffic identification," Blackhat *USA*, 2015.

21. M. Usha and P. Kavitha, "Anomaly based intrusion detection for 802.11 networks with optimal features using svm classifier," *Wireless Networks*, pp. 1–16, Springer, 2016.
22. H Nguyen, K Franke, S Petrovic Improving Effectiveness of Intrusion Detection by Correlation Feature Selection, 2010 International Conference on Availability, Reliability and Security,IEEE Pages-17-24
23. S Chebrolu, A Abraham, J P. Thomas Feature deduction and ensemble design of intrusion detection systems, Computers & Security, Vol. 24, No. 4, pp. 295-307, 2005.
24. T. S. Chou, K. K. Yen, and J. Luo "Network Intrusion Detection Design Using Feature Selection of Soft Computing Paradigms. International Journal of Computational Intelligence 4;3 2008
25. S. Axelsson, "The base rate fallacy and its implications for the difficulty of Intrusion detection", Proc. Of 6th. ACM conference on computer and communication security 1999.
26. R. Puttini, Z.marrakchi, and L. Me, "Bayesian classification model for Real time intrusion detection", Proc. of 22nd. International workshop on Bayesian inference and maximum entropy methods in science and engineering, 2002.
27. https://scikitlearn.org/stable/modules/generated/sklearn.feature_selection.SelectFromModel.html#sklearn.feature_selection.SelectFromModel
28. Razan Abdulhammed, Hassan Musafer , Ali Alessa , Miad Faezipour and Abdelshakour Abuzneid "Features Dimensionality Reduction Approaches for Machine Learning Based Network Intrusion Detection", MDPI journal on Electronics, Vol. 8, pp.322-329, 2019.
29. Nivaashini.M, Thangaraj.P "State-Of-The-Art Machine Learning and Deep Learning: Evolution of Intelligent Intrusion Detection System against Wireless Network (WiFi) Attacks in Internet of Things (Iot)" International Journal of Innovative Technology and Exploring Engineering (IJITEE), Vol. 8, No.3, 2019.
30. Vu Viet Thang and F. F. Pashchenko "Multistage System-Based Machine Learning Techniques for Intrusion Detection in WiFi Network", Journal of Computer Networks and Communications, Vol. 2019, Article ID 4708201, pp. 1-13, 2019.
31. Markus Ring, Sarah Wunderlich, Deniz Scheuring, Dieter Landes and Andreas Hotho" A Survey of Network-based Intrusion Detection Data Sets", pp. 1-17, 2019. arXiv:1903.02460 [cs.CR]
32. Feng, G., Li, B., Yang, M., & Yan, Z. "V-CNN: Data Visualizing based Convolutional Neural Network", 2018 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC), pp. 1-7, 2018.
33. Qin, Y., Li, B., Yang, M., & Yan, Z, "Attack Detection for Wireless Enterprise Network: A Machine Learning Approach", IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC), 2018.
34. Vaca, F.D., &Niyaz,Q, "An Ensemble learning based Wi-Fi Network Intrusion Detection system (WNIDS)", IEEE 17th International Symposium on Network Computing and Applications (NCA), 2018.
35. S Ganapathy, P Yogesh, A Kannan, "Intelligent agent-based intrusion detection system using enhanced multiclass SVM", Computational intelligence and neuroscience, Vol. 2012, pp. 1-9, 2012.
36. S Ganapathy, K Kulothungan, S Muthurajkumar, M Vijayalakshmi, P.Yogesh, A.Kannan, "Intelligent feature selection and classification techniques for intrusion detection in networks: a survey", EURASIP Journal on Wireless Communications and Networking, Vol.271, No.1, pp. 1-16, 2013.
37. A.R Rajeswari, K Kulothungan, S Ganapathy, A Kannan, "Trust aware SVM based IDS for Mitigating the Malicious Nodes in MANET", International Journal of Innovative Technology and Exploring Engineering, Vol. 8, No.8, pp. 185-197, 2019.
38. M Selvi, K Thangaramya, Sannasi Ganapathy, Kanagasabai Kulothungan, H Khannah Nehemiah, Arputharaj Kannan, "An Energy Aware Trust Based Secure Routing Algorithm for Effective Communication in Wireless Sensor Networks", Wireless Personal Communications, Vol. 105, No. 4, pp. 1475-1490, 2019.
39. AR Rajeswari, K Kulothungan, Sannasi Ganapathy, A Kannan, "A trusted fuzzy based stable and secure routing algorithm for effective communication in mobile adhoc networks", Peer-to-Peer Networking and Applications, pp. 1-21, 2019. Article in Online:
40. S Muthurajkumar, S Ganapathy, M Vijayalakshmi, A Kannan, "An Intelligent Secured and Energy Efficient Routing Algorithm for MANETs", Wireless Personal Communications, Vol. 96, No.2, pp. 1753-1769, 2018.
41. U Kanimozhi, S Ganapathy, D Manjula, A Kannan, "An Intelligent Risk Prediction System for Breast Cancer Using Fuzzy Temporal Rules", National Academy Science Letters, Vol.42, No.03, pp. 227-232, 2019.
42. D Sudaroli Vijayakumar, S Ganapathy, "Machine Learning Approach to Combat False Alarms in Wireless Intrusion Detection System", Computer and Information Science, Vol.11, No.3, pp. 67-81, 2018.
43. R Logambigai, Sannasi Ganapathy, Arputharaj Kannan, "Energy–efficient grid–based routing algorithm using intelligent fuzzy rules for wireless sensor networks", Computers & Electrical Engineering, Vol.68, pp. 62-75, 2018.
44. B PrabhuKavin, S Ganapathy, "Data Mining Techniques for Providing Network Security through Intrusion Detection Systems: A Survey", International Journal of Advances in Computer and Electronics Engineering, Vol.2, No.10, pp. 1-6, 2017.
45. S Ganapathy, N Jaisankar, P Yogesh, A Kannan, "An intelligent system for intrusion detection using outlier detection", 2011 International Conference on Recent Trends in Information Technology (ICRTIT), pp. 119-123, 2011

AUTHORS PROFILE

D. Sudaroli Vijayakumar, currently pursuing her doctorate in VIT holds master's in digital communication and Networking. Her research interests are machine learning and allied areas. Her achievements include Let's Clone you, Fast tracker awards. She is a member of CSI and IAENG.