Intrusion Detection System Using Feature Selection Approach

[1]K.Spandana, [2]M.Bhanu Chandrika

[1]CSE Department, [2] CSE Department
CBIT, Hyferabad
Email - spandana.k@cbit.ac.in, bhanu.mudundi@gmail.com

Abstract: *A tremendous growth in the usage of internet services has raised various concerns pertaining to the protection of internet applications and computer networks against threats from ever evolving cyber attacks. Therefore, the development of effective and adaptive security approaches has become extremely important, since traditional security techniques such as user authentication, firewalls and data encryption, are insufficient to fully cover the entire landscape of network security. Hence, another line of security defense is highly recommended, such as Intrusion Detection System (IDS), which is s a device or software application that monitors a network or systems for malicious activities or policy violations. Any such detected activity or violation is typically either reported to an administrator or collected centrally using a security information and event management system. This project aims to introduce an Intrusion Detection System (IDS) named Least Square Support Vector Machine based IDS (LSSVMIDS), that will be constructed using the features selected by the proposed feature selection algorithm which will analytically select the optimal feature for classification of patterns that do not match normal network traffic.*

Key Words: *Intrusion Detection System (IDS), Least Square Support Vector Machine based IDS (LSSVMIDS), Cyber Attacks*

1. INTRODUCTION:

Network traffic classification is often impacted with certain long-term problems such as extraneous and unessential features present in data, which not only slow down the process of classification but also prevent a classifier from making well-informed and accurate decisions, which are free from bias, especially when coping with vast amounts of data. An unsupervised mutual information-based approach that can be used to rationally determine the most favourable features for classification is proposed in order to aid the classification and intrusion detection processes. This mutual information-based feature selection algorithm can handle linearly and nonlinearly dependent data features, whose effectiveness can be studied and evaluated in the cases of network intrusion detection. A Least Square Support Vector Machine based IDS (LSSVM-IDS), is used for this purpose. The IDS is built using the features selected by the proposed feature selection approach and its performance will be analyzed using a standard intrusion detection evaluation dataset, namely the KDD Cup 99 dataset.

2. LITERATURE REVIEW:

Feature selection is a technique for eliminating irrelevant and redundant features in order to obtain most optimal subset of features that produce a better characterization of patterns belonging to different classes. Methods for feature selection are generally classified into filter and wrapper methods [1]. Filter algorithms utilize an independent measure (such as, information measures, distance measures, or consistency measures) as a criterion for estimating the relation of a set of features, while wrapper algorithms make use of particular learning algorithms to evaluate the value of features. In comparison with filter methods, wrapper methods are often much more computationally expensive when dealing with high dimensional data or large scale data. In this study hence, we focus on filter methods for IDS. Due to the continuous growth of data dimensionality, feature selection as a preprocessing step is becoming an essential part in building intrusion detection systems [3]. Recently, Ambusaidi and Nanda [2] proposed a forward feature selection algorithm using the mutual information method to measure the relation among features. The optimal feature set was then used to train the LS-SVM classifier and to build the IDS. Pervez and Farid [3] presented an intrusion classification approach based on the combination of feature selection and SVM classifier, which achieved 99% accuracy using three features.

3. METHODOLOGY:

The architecture of the proposed system contains the following modules as shown in the figure 1.
1. Data collection 2• Data Preprocessing 3• Filter based feature selection 4 • Attack classification & Recognition 5• Performance Evaluation

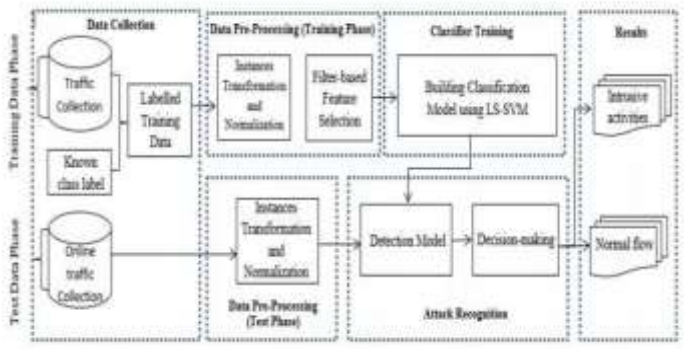


FIGURE 1:ARCHITECTURE OF LS-SVM-BASED INTRUSION DETECTION SYSTEM [1]

The above figure illustrates the architecture of LS-SVM-based Intrusion Detection System which consists of the following stages: Data collection, Data Preprocessing, Filter based feature selection, Attack classification & Recognition and Performance Evaluation.

Algorithm 1

Algorithm for Flexible Mutual Information Based Feature Selection

Input: Feature set F = {fi,i = 1,...,n}

Output: S - the selected feature subset begin

Step 1. Initialization: set S = φ 11

Step 2. Calculate I(C;fi) for each feature, i = 1,...,n

Step 3.nf = n; Select the feature fi such that: argmax fi (I(C;fi)),i = 1,...,nf, Then, set F ← F\{fi }; S ←S∪{fi }; nf = nf−1.

Step 4. while F≠ φ do Calculate GMI in (1) to find fi where i∈{1,2,⋯ ,nf}; nf = nf −1; F ← F\{fi}; if (GMI > 0) then S ←S ∪{fi }.

end

end

Step 5. Sort S according to the value of GMI of each selected feature.

return S

Algorithm 2

Algorithm for Feature Selection Based On Linear Correlation Coefficient

Input: Feature set F = {fi,i = 1,...,n}

Output: S - the selected feature subset Begin

Step 1. Initialization: S = φ

Step 2. Calculate corr(C;fi) for each feature, i = 1,...,n

Step 3.nf = n; Select the feature fi such that: argmax fi (corr(C;fi)),i = 1,...,nf, Then, set F ← F\{fi }; S ←S∪{fi }; nf = nf−1.

Step 4. while F≠φ do Calculate Gcorr in (4) to find fi where i∈{1,2,⋯ ,nf}; nf = nf −1; F ← F\{fi}; if (Gcorr> 0) then S ←S ∪{fi }. endend 13

Step 5. Sort S according to the value of Gcorr of each selected feature.

return S

Algorithm 3

Algorithm For Intrusion Detection Based On LS-SVM

Input: LS-SVM Normal Classifier, selected features (normal class), an observed data item x

Output: Lx - the classification label of x

Begin

Lx ←classification of x with LS-SVM of Normal class

if Lx == “Normal” then

Return LX

else

do: Run Algorithm 4 to determine the class of attack

end

end

Algorithm 4

Algorithm For Attack Classification Based On LS-SVM
Input: LS-SVM Normal Classifier, selected features (normal class), an observed data item x
Output: Lx - the classification label of x

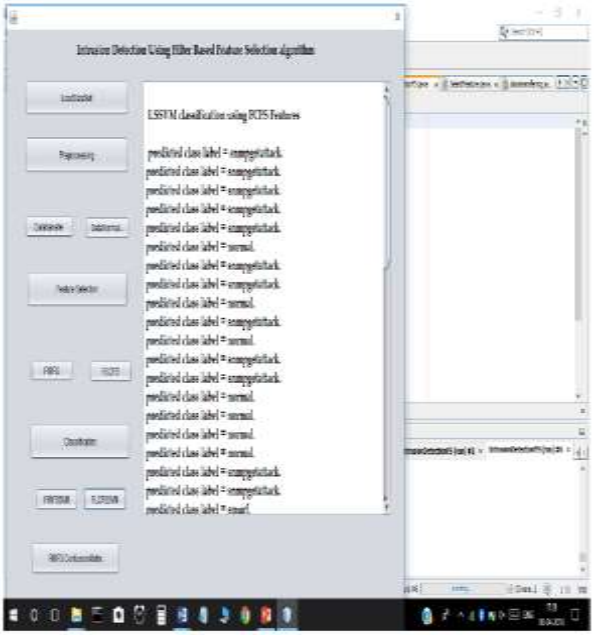
```
Begin
    Lx ←classification of x with LS-SVM of DoS class
    if Lx==“DoS” then
        Return LX
    else
    Lx ←classification of x with LS-SVM of Probe class
    if Lx == “Probe” then
 Return LX
else
 Lx ←classification of x with LS-SVM of R2L class
if Lx == “R2L” then
Return LX
else
 Lx == “U2R”; Return LX
End
End
End
```

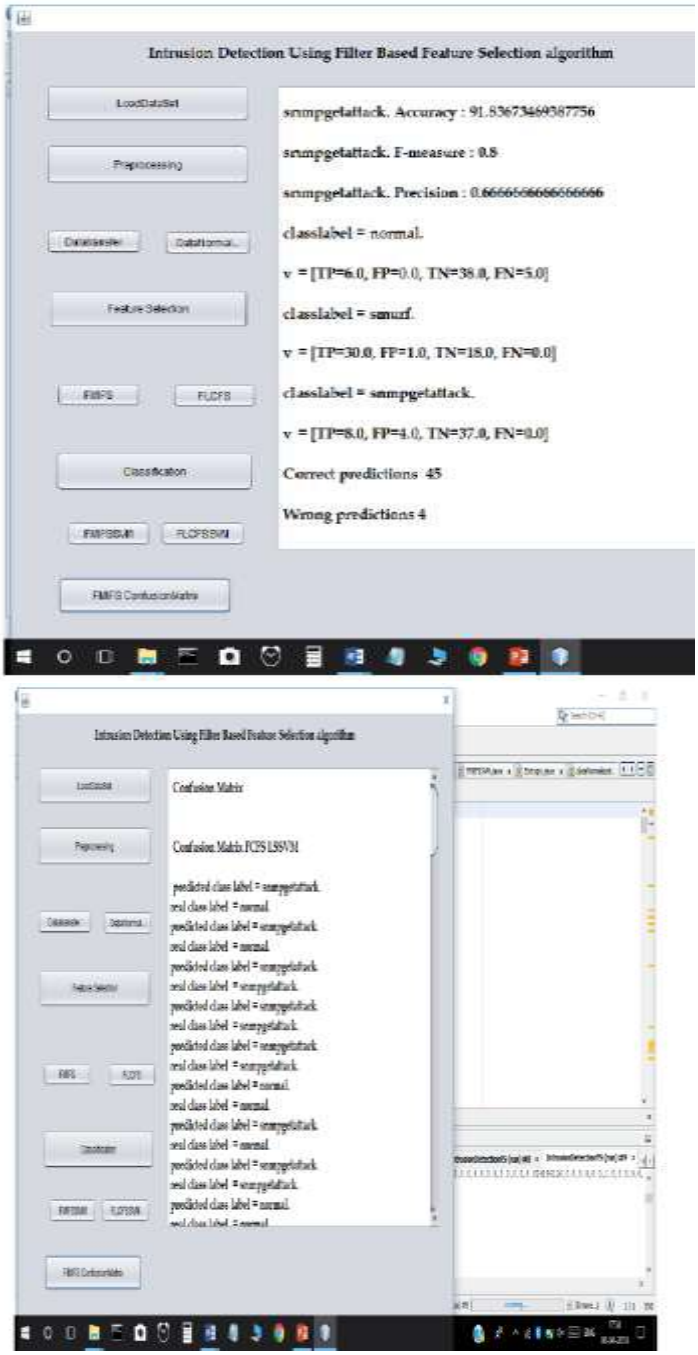
4. RESULTS AND DISCUSSION:

A hybrid intrusion detection system based on data mining concepts has been designed for classification and detection of intrusive events. The major contributions provided are (i) Effective intrusion detection, (ii)The detection techniques for network and/or host intrusion detection systems that use classification algorithms to enhance the performance of the intrusion detection system. The proposed system can also be further enhanced by optimizing the search strategy. Also the performance is to be evaluated using In order to evaluate the performance and effectiveness of the proposed LSSVMIDS, the Accuracy, Detection Rate, False Positive Rate, F-measure, Precision and Recall metrics are applied:

Accuracy = (TP + TN) / (TP + TN + FN + FP)
Detection Rate = (TP)/ (TP + FN)
False Positive Rate =(FP)/ (FP + TN)
Precision=TP/(TP+FP)
Recall=TP/(TP+FN)
F-measure=2(Precision*Recall)/(Precision+Recall)

where, True Positive (TP) is the number of actual attacks classified as attacks, True Negative (TN) is the number of actual normal records classified as normal ones, False Positive (FP) is the number of actual normal records classified as attacks, and False Negative (FN) is the number of actual attacks classified as normal records.





5. CONCLUSION AND FUTURE WORK:

A robust classification method in conjunction with an efficient feature selection algorithm is the main components required to build a well-functioning Intrusion Detection System. In this project, a supervised filter-based feature selection algorithm, namely Flexible Mutual Information Feature Selection (FMIFS) is done, which is an overall improvement over MIFS and MMIFS. FMIFS suggests a modification to Battiti's algorithm to reduce the redundancy among features by eliminating the b parameter, which has been used with no specific procedure or guidelines. The system uses Support Vector Machine methods, namely Least Square SVMs in order to build the IDS. LSSVM is a least square version of SVM that works with equality constraints instead of inequality constraints in the formulation designed to solve a set of linear equations for classification problems rather than a quadratic programming problem. The proposed LSSVM-IDS + FMIFS has been evaluated using the KDD Cup 99 dataset. Although the feature selection algorithm has shown encouraging performance, it could be further enhanced by optimizing the search strategy. In addition, adoptive learning algorithms can be used to impact the unbalanced sample distribution on an IDS. In the future, other mining classifiers can be ensembled with Support Vector Machines to achieve better classification accuracy for the minority class instances. In the future, other mining classifiers can be ensemble with Support Vector Machines to achieve better classification accuracy for the minority class instances.

REFERENCES:

1. Mohammed A. Ambusaidi, Xiangjian He, Priyadarsi Nanda, Zhiyuan Tan, Building 9an intrusion detection system using a filter-based feature selection algorithm, IEEE, 2016.
2. Y. Chang, WeiLi, Z.Yang, Network Intrusion detection based on random forest and support vector machine, IEEE, 2017.
3. A.M. Ambusaidi, X. He, Z. Tan, P. Nanda, Unsupervised feature selection method for intrusion detection system, IEEE, 2016.
4. H. Gharaee, Hamid, A new feature selection IDS based on genetic algorithm and SVM, IEEE, 2016.
5. Praneeth NSKH, Naveen Varma, Principle component analysis based intrusion detection system using support vector machine, IEEE, 2016.
6. A.M. Ambusaidi, X. He, Z. Tan, P. Nanda, A novel feature selection approach for intrusion detection data classification, IEEE,2016.
7. R. Chitrakar and C. Huang, Selection of candidate support vectors in incremental SVM for network intrusion detection, IEEE, 2015.
8. M.S. Pervez, Dewan Md. Farid, Feature selection and intrusion classification in NSLKDD cup 99 Dataset employing SVM, IEEE, 2014
9. S. Cang and H. Yu., Mutual information based input feature selection for classification problems,IEEE,2014
10. M. A. Salama, H. F. Eid, R. A. Ramadan, A. Darwish, and A. E. Hassanien, Hybrid intelligent intrusion detection scheme, IEEE, 2014.