International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

A PROJECT REPORT ON INTRUSION DETECTION

MOHD AZHAR SIDDIQUI[1], RAHMATH ALI[2]

1. Student of Department of computer science and engineering & ISL Engineering College
2. Assistant Professor of Department of computer science and engineering & ISL Engineering College

ABSTRACT

Intrusion detection is a classification task, and it consists of building a predictive model which can identify attack instances. There are too many features or attributes which may contain false correlation. Moreover, many features may be irrelevant or redundant. For this reason, feature selection methods can be used to get rid of the irrelevant and redundant features without decreasing performance. A mutual information based algorithm analytically selects the optimal feature for classification and can handle linearly and nonlinearly dependent data features. An Intrusion Detection System (IDS), named Least Square Support Vector Machine based IDS (LSSVM-IDS), is built using the features selected by feature selection algorithm. As a well-known intrusion evaluation dataset, KDD Cup 99 dataset is a typical example of large-scale datasets. The KDD Cup 99 datasetis one of the most popular and comprehensive intrusion detection datasets and is widely applied to evaluate the performance of intrusion detection systems. A general problem regarding public datasets is sparse data, which is a recognized challenge when working with machine learning within classifiers. Some time in data mining some source of data generators will generate error or unrelated data and this data will put effect when performing classification. This unrelated data we can remove using sparse matrix concept which compare similarity between two records, if two records are completely unrelated then there will be less similarity and it will be remove out. Using this technique we can reduce dataset size and can improve classification.

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

INTRODUCTION

In recent years, network attacks and information security incidents occurred frequently, covering areas more and more widely, and increasingly harmful. Network information securityhas become more and more serious as the rapid development of the computer network. Despite increasing realization of network security, the existing solutions remain incompetent of fully protecting internet applications and computer networks against the intimidation from ever- advancing cyber attack techniques such as DoS attack and computer malware. Developing efficient and adaptive security approach, therefore, has become more serious than ever before. The established security technique, as the first line of security defence, such as user verification, firewall and data encryption, are unsatisfactory to fully cover the entire landscape of network security while opposite challenge from ever-evolving intrusion skills and techniques. Hence, one more line of security defence is greatly suggested, such as Intrusion Detection System (IDS). Recently, an IDS next to with anti-virus software has become an important complement to the security infrastructure of most organizations. The blend of these two lines provides a more complete defence against those threats and enhances network security. An intrusion detection system monitors network traffic for suspicious activity and alerts thesystem or network administrator in order to take evasive action. It has a very important position in the network information security and it is considered as the second security gate after firewall. Intrusion detection is a classification task, and it consists of constructing a predictive model which can identify attack instances. On the one hand, there are too many features or attributes which may contain false correlation. Moreover, many features may be irrelevant or redundant.

Due this reason, feature selection methods can be used to get purge of the irrelevant and surplus features without decreasing act.

AIM OF THE PROJECT

The main aim is to build an intrusion detection system which combines a simple feature selection algorithm and SVM technique to see attacks. Using KDD cup data set and Data Mining take out the hidden prognostic information from large Databases. It is a influential new technology with great potential that helps companies focus on the most important information in their data warehouses

MOTIVATION

The motivation for doing this project was primarily interest in undertaking challenging projectin an interesting area of providing network security in data mining by building an IDS using machine learning technique. An IDS monitor network traffic for doubtful activity and alerts the system or network administrator in order to take evasive action.

APPLICATION OF THESIS

This is a type of security management system for computers and networks. An IDS system gathers and analyzes information from various areas within a computer or a network to identify possible security breaches, which include both intrusions (attacks from outside the organization) and misuse (attacks from within the organization). ID uses vulnerability assessment (sometimes referred to as scanning), which is a

technology developed to assessthe security of a computer system or network.

This thesis can be applicable for:

- Monitoring and analyzing both user and system activities
- Analyzing system configurations and vulnerabilities Assessing system and file integrity
- Ability to recognize patterns typical of attacks
- Analysis of abnormal activity patterns
- Tracking user policy violations

The purpose of IDS is to help computer systems on how to deal with attacks, and that IDS is collecting information from several different sources within the computer systems and networks and compares this information with pre-Existing patterns of discrimination as to whether there are attacks or weaknesses.

SCOPE OF THESIS

The fundamental scope of the project is to propose a mutual information based algorithm that analytically selects the optimal feature for classification. This mutual information based feature selection algorithm can handle linearly and nonlinearly dependent data features.

Its effectiveness is evaluated in the cases of network intrusion detection. An Intrusion Detection System (IDS), named Least Square Support Vector Machine based IDS (LSSVM-IDS), is built using the features selected by our proposed feature selection algorithm. The performance of LSSVM-IDS is evaluated using three intrusion detection evaluation datasets, namely KDD Cup 99, NSL-KDD and Kyoto 2006+ dataset. The evaluation results show that our feature selection algorithm contributes more criticalfeatures for LSSVM-IDS to achieve better accuracy and lower computational cost.

ORGANISATION OF THESIS

The thesis is organized into seven chapters. Chapter 1 gives the brief introduction to the thesis. Chapter 2 describes Literature Survey,, Chapter 3 is about System analysis; it gives the brief idea about the existing system and its problems, proposed system and its detail description. Chapter 4 is Algorithm and Techniques, explains all the algorithms andtechniques used in the project.

Chapter 5 describes experimental results and Evaluation. Chapter 6 describes Testing and Test cases Finally, Chapter 7 provides conclusion and future development of the project.

LITERATURE SURVEY

Title: Building an intrusion detection system using filter based feature selection algorithm

AUTHORS: Mohammed A. Ambusaidi, Xiangjian He, Priyadarsi Nanda, Zhiyuan Tan Redundant and irrelevant features in data have caused a long-term problem in network traffic classification. These features not only slow down the process of classification but also prevent a classifier from making accurate decisions, especially when coping with big data. In this paper, A mutual information based algorithm was proposed that analytically selects the optimal feature for classification. This mutual information based feature selection algorithm can handle linearly and nonlinearly dependent data features. Its effectiveness is evaluated in the cases of network intrusion detection. An Intrusion Detection System (IDS), named Least SquareSupport Vector Machine based IDS

(LSSVM-IDS), is built using the features selected by proposed feature selection algorithm. The performance of LSSVM-IDS is evaluated using three intrusion detection evaluation datasets, namely KDD Cup 99, NSLKDD and Kyoto 2006+ dataset. The evaluation results show that feature selection algorithm contributes more critical features for LSSVM-IDS to achieve better accuracy and lower computational cost compared with the state-of-the-art methods. Index Terms— Intrusion detection, Feature selection, Mutual information, Linear correlation coefficient, Least square support vector machine

Title: Traffic-aware design of a high speed FPGA network intrusiondetection system

AUTHORS: S. Pontarelli, G. Bianchi, S. Teofili Security of today's networks heavily rely on network intrusion detection systems (NIDSs). The ability to promptly update the supported rule sets and detect new emerging attacks makes field-programmable gate arrays (FPGAs) a very appealing technology. An important issue is how to scale FPGA-based NIDS implementations to ever faster network links. Whereas a trivial approach is to balance traffic over multiple, but functionally equivalent, hardware blocks, each implementing the whole rule set (several thousands rules), the obvious cons is the linear increase in the resource occupation. In this work, the promotion of different, traffic- aware, modular approach in the design of FPGA-based NIDS is done. Instead of purely splitting traffic across equivalent modules, classifying and group homogeneous traffic, and dispatch it to differently capable hardware blocks, each supporting a (smaller) rule set tailored to the specific traffic category. Implementation and validation is performed using the rule set of the well-known Snort NIDS, and experimentally investigate the emerging trade-offs and advantages, showing resource savings up to 80 percent based on real-world traffic statistics gathered from an operator's backbone.

Title: An effective technique for intrusion detection using neurofuzzyand radial SVM classifier.

AUTHORS: A. Chandrasekhar, K. Raghuveer Intrusion detection is not yet a perfect technology. This has given data mining the opportunity to make several important contributions to the field of intrusion detection. In this paper, They have proposed a new technique by utilizing data mining techniques such as neuro- fuzzy and radial basis Support Vector Machine (SVM) for the intrusion detection system. The
proposed technique has four major steps in which, first step is to perform the Fuzzy C-means clustering (FCM). Then, neuro-fuzzy is trained, such that each of the data point is trained with the corresponding neuro-fuzzy classifier associated with the cluster. Subsequently, a vector for SVM classification is formed and in the fourth step, classification using radial SVM is performed to detect intrusion has happened or not. Data set used is the KDD cup 99 dataset and they have used sensitivity, specificity and accuracy as the evaluation metrics parameters. Their

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

technique could achieve better accuracy for all types of intrusions. It achieved about 98.94 % accuracy in case of DOS attack and reached heights of 97.11 % accuracy in case of PROBE attack. In case of R2L and U2R attacks it has attained 97.78 and 97.80 % accuracy respectively. Compared the proposed technique with the other existing state of art techniques. These comparisons proved the effectiveness of our technique.

SYSTEM ANALYSIS AND DESIGN OVERVIEW

The traditional security techniques, as the first line of security defence, such as user authentication, firewall and data encryption, are insufficient to fully cover the entire landscape of network security while facing challenges from ever-evolving intrusion skills and techniques
. Hence, another line of security defence is highly recommended, such as Intrusion Detection System (IDS). Recently, an IDS alongside with anti-virus software has become an important complement to the security infrastructure of most organizations.

Machine learning techniques, such as Support Vector Machine (SVM), to classify networktraffic patterns that do not match normal network traffic. Both systems were equipped with five distinct classifiers to detect normal traffic and four different types of attacks (i.e., DoS, probing, U2R and R2L). Experimental results show the effectiveness and robustness of using SVM in IDS.

However, current network traffic data, which are often huge in size, present a major challenge to IDSs. These ―big data‖ slow down the entire detection process and may lead to unsatisfactory classification accuracy due to the computational difficulties in handling such data. Classifying a huge amount of data usually causes many mathematical difficulties which then lead to higher computational complexity. As a well-known intrusion evaluation dataset, KDD Cup 99 dataset is a typical example of large-scale datasets. This dataset consists of more than five million of training samples and two million of testing samples respectively. Such a large scale dataset retards the building and testing processes of a classifier, or makes the classifier unable to perform due to system failures caused by insufficient memory. To address the aforementioned problems on the methods for feature selection, a hybrid feature selection algorithm was proposed (HFSA). HFSA consists of two phases. The upper phase conducts a preliminary search to eliminate irrelevant and redundancy features from the original data. This helps the wrapper method (the lower phase) to decrease the searching rangefrom the entire original feature space to the pre-selected features (the output of the upper phase). This work proposes a new filter-based feature selection method, in which theoretical analysis of mutual information is introduced to evaluate the dependence between features and output classes. The most relevant features are retained and used to construct classifiers forrespective classes.

Conducts complete experiments on two well known IDS datasets in addition to the datasetused.

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

This is very important in evaluating the performance of IDS since KDD dataset is outdated and does not contain most novel attack patterns in it. In addition, these datasets are frequently used in the literature to evaluate the performance of IDS. Moreover, these datasets have various sample sizes and different numbers of features, so they provide a lot more challenges for comprehensively testing feature selection algorithms. Different from the detection framework proposed that designs only for binary classification, we design our proposed framework to consider multiclass classification problems. This is to show the effectiveness and the feasibility of the proposed method.

EXISTING SYSTEM

A significant amount of research has been conducted to develop intelligent intrusion detection techniques, which help achieve better network security. Bagged boostingbased on C5 decision trees and Kernel Miner are two of the earliest attempts to build intrusion detection schemes.

Mukkamala et al. investigated the possibility of assembling various learning methods,

including Artificial Neural Networks (ANN), SVMs and Multivariate Adaptive Regression Splines (MARS) to detect intrusions.

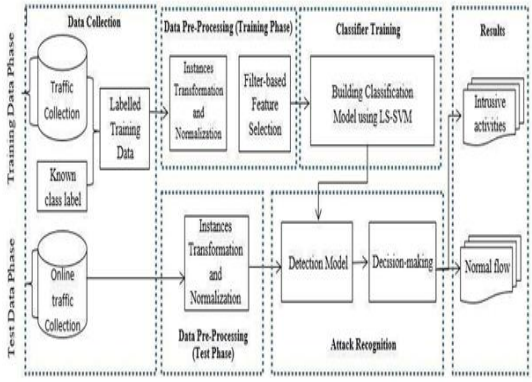


Fig:3.1 Existing System Architecture

DISADVANTAGES:

 Existing solutions remain incapable of fully protecting internet applications and computer networks against the threats from ever-advancing cyber attack techniques such as DoS attack and computer malware.

 Current network traffic data, which are often huge in size, present a major challenge to IDSs. These ―big data‖ slow down the entire detection process and may lead to unsatisfactory classification accuracy due to the computational difficulties inhandling such data.

 Classifying a huge amount of data usually causes many mathematical difficulties which then lead to higher computational complexity.

Largewhich present critical challenges to knowledge discovery and data modelling.-scale datasets usually contain noisy, redundant, or uninformative features

Sparsity causes the major problem as data is huge in size.

PROBLEM STATEMENT:

Imagine that you have a data with lots of null or impossible values. These valuesrepresents the sparsity in your data.

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

If these values are covering big part of your table, they can be represented as a sparsematrix to save some memory.

A common problem in machine learning is sparse data, which alters the performance of machine learning algorithms and their ability to calculate accurate predictions.

Data is considered sparse when certain expected values in a dataset are missing, which is a common phenomenon in general large scaled data analysis.

A general problem regarding public datasets is sparse data, which is a recognized challenge when working with machine learning within classifiers.

Sparse data refers to data that is incomplete and can have an immense effect on the ability to train the classifier into producing accurate predictions

PROPOSED SYSTEM

The proposed Hybrid Feature Selection Algorithm (HFSA) consists of two phases.

The upper phase conducts a preliminary search to eliminate irrelevant and redundancyfeatures from the original data. This helps the wrapper method (the lower phase)
to decrease the searching range from the entire original feature space to the pre-selected features (the output of the upper phase). The key contributions of this paper are listed as follows.

This work proposes a new filter-based feature selection method, in which theoretical analysis of mutual information is introduced to evaluate the dependence bet weenfeatures and output classes.

The most relevant features are retained and used to construct classifiers for respective classes. As an enhancement of Mutual Information Feature Selection (MIFS) and Modified Mutual Information based Feature Selection (MMIFS), the proposed featureselection method does not have any free parameter, such as in MIFS and MMIFS. Therefore, its performance is free from being influenced by any inappropriate assignment of value to a free parameter and can be guaranteed. Moreover, the proposed method is feasible to work in various domains, and more efficient in comparison with HFSA, where the computationally expensive wrapper-based feature
selection mechanism is used.

Also conducted complete experiments on two well known IDS datasets in addition to the dataset used. This is very important in evaluating the performance of IDS since KDD dataset is outdated and does not contain most novel attack patterns in it. In addition, these datasets are frequently used in the literature to evaluate the performance of IDS. Moreover, these datasets have various sample sizes and different numbers of features, so they provide a lot more challenges for comprehensively testing feature selection algorithms.

Different from the detection framework proposed that designs only for binary classification, we design our proposed framework to consider multiclass classification problems. This is to show the effectiveness and the feasibility of the proposedmethod.

In this project I added sparse matrix concept using which we can remove erroneous and unrelated data.

Some time in data mining some source of data generators (sensors sensing

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

temperature
or IDS system which will record network logs) will generate error or

 This unrelated data we can remove using sparse matrix concept which compare similarity between two records, if two records are completely unrelated then there
will be less similarity and it will be remove out.

Using this technique we can reduce dataset size and can improve classification.

ADVANTAGES:

FMIFS is an improvement over MIFS and MMIFS.

FMIFS suggests a modification to Battiti's algorithm to reduce the redundancyamong features. FMIFS eliminates the redundancy parameter required in
MIFS and MMIFS.

Using thistechnique we can reduce dataset size and can improve classification

PROPOSED SYSTEM ARCHITECTURE

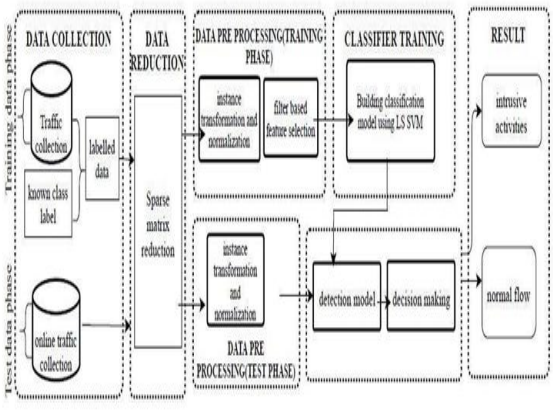


Fig: 3.2 Proposed System Architecture

The framework of the proposed intrusion detection system is depicted in Figure. The detection framework is comprised of four main phases: (1) data collection, where sequences of network packets are collected, (2) data pre processing, where training and test data are pre processed and important features that can distinguish one class from the others are selected,
(3) classifier training, where the model for classification is trained using LS-SVM, and (4) attack recognition, where the trained classifier is used to detect intrusions on the test data. The following subsections explain each phase in detail.

DATA COLLECTION: Data collection is the first and a critical step to intrusion detection. The type of data source and the location where data is collected from are two determinate factors in the design and the effectiveness of an IDS. To provide the best suited protection for the targeted host or networks, this study proposes a network- based IDS to test our proposed approaches. The proposed IDS runs on the nearest
router to the victim(s) and monitors the inbound network traffic.

During the training stage, the collected data samples are categorised with respect to the transport/Internet layer protocols and are labelled against the domain knowledge. However, the data collected in the test stage are categorized according to the protocol types only.

DATA REDUCTION: Sparse matrix concept is used using which we can remove erroneous and unrelated data. Some time in data mining some source of data generators (sensors sensing temperature or IDS system which will record network logs) will generate error or unrelated data and this data will put effect when performing classification.

This unrelated data we can remove using sparse matrix concept which

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

compare similarity between two records, if two records are completely unrelated then there will be less similarity and it will be remove out. Using this technique we can reduce dataset size and can improve classification.

 DATA PRE PROCESSING: The data obtained during the phase of data collection

are first processed to generate the basic features such as the ones in KDD Cup 99 dataset. This phase contains three main stages shown as follows.

Data Transferring: The trained classifier requires each record in the input data to be represented as a vector of real number. Thus, every symbolic feature in a dataset is first converted into a numerical value.

Data Normalisation: An essential step of data pre processing after transferring all symbolic attributes into numerical values is normalisation. Data normalisation is a process of scaling the value of each attribute into a well proportioned range, so that the bias in favour of features with greater values is eliminated from the dataset. Every feature within each record is normalised by the respective maximum value and falls into the same range of [0-1]. The transferring and normalisation process will also be applied to test data.

For KDD Cup 99 and to make a comparison with those systems that have beenevaluated on different types of attacks and five classes are constructed. One of these classes contains purely the normal records and the other four hold different types of attacks (i.e., DoS, Probe, U2R, R2L), respectively.

Feature Selection: Even though every connection in a dataset is represented byvarious features, not all of these features are needed to build an IDS. Therefore, it is important to identify the most informative features of traffic data to achieve higher performance. However, the proposed feature selection algorithms can only rank features in terms of their relevance but they cannot reveal the best number of features that are needed to train a classifier. Therefore, this study applies the same technique proposed to determine the optimal number of required features. To do so, the technique first utilizes the proposed feature selection algorithm to rank all features based on their importance to the classification processes. Then, incrementally the technique adds features to the classifier one by one. The final decision of the optimal number of features in each method is taken once the highest classification accuracy in the training dataset is achieved.

CLASSIFIER TRAINING: Once the optimal subset of features is selected, this subset is then taken into the classifier training phase where LS-SVM is employed. Since SVMs can only handle binary classification problems and because for KDD Cup 99 five optimal feature subsets are selected for all classes, five LS-SVM classifiers need to be employed. Each classifier distinguishes one class of records from the others. For example the classifier of Normal class distinguishes Normal data from non-Normal(All types of attacks). The DoS class distinguishes DoS traffic from non-DoS data (including Normal, Probe, R2L and U2R instances) and so on. The five LS-SVM classifiers are then combined to build the intrusion detection model to distinguish all different classes. class are reported as normal data,

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
ISSN: 2457-0362

otherwise are considered as attacks. After completing all the aforementioned steps and the classifier is trained using the optimal subset of features which includes the most correlated and important features, the normal and intrusion traffics can be identified by using the saved trained

ATTACK RECOGNITION: In general, it is simpler to build a classifier to distinguish between two classes than considering multi classes in a problem. This is because the decision boundaries in the first case can besimpler. The first part of the experiments in this paper uses two classes, where records matching to the normal classifier. The test data is then directed to the saved trained model to detect intrusions. Records matching to the normal class are considered as normal data, and the other records are reported as attacks. If the classifier model confirms that the record is abnormal, the subclass of the abnormal record (type of attacks) can be used to determine the record's type.

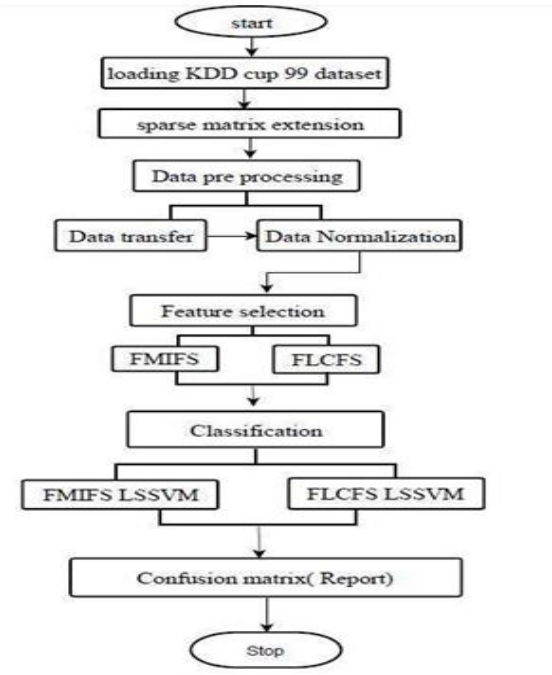
MODULE DESCRIPTION:

MODULES:

1. Data Pre processing
2. Sparse Matrix extension
3. Filter based feature selection
4. Attack classification & Recognition
5. Performance Evaluation

FLOW CHART

FIG 3.3 FLOW **CHART** 4.



ALGORITHMS AND TECHNIQUES

An algorithm is a procedure or formula for solving a problem, based on executing a sequence of specified actions. A computer program can be viewed as an elaborate algorithm. In mathematics and computer science, an algorithm usually means a small procedure that solves a recurrent problem.

EXPERIMENTAL RESULTS AND EVALUATION

This chapter shows the implementation details of proposed system project. It also shows the steps required to generate the report for the intrusion detection system. Also, it introduces the project testing using different testing environments and the results after testing. It also introduces a GUI we implement to facilitate interaction with the system and make it very comfortable.

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

RESULT SCREENSHOTS

A screenshot (or screen grab) is a digital image of what should be visible on a monitor, television, or other visual output device. Screenshots clearly explain the project through Graphical User Interface(GUI). They are the front end pages of the implementation. They represent the entire flow of application.

1. **Home page:**

Fig 5.1 Main Home page

Loading data set: The 1st step is to load the DD cup 99 data set. Clickon the Load Dataset button to upload the data. This shows the data size.

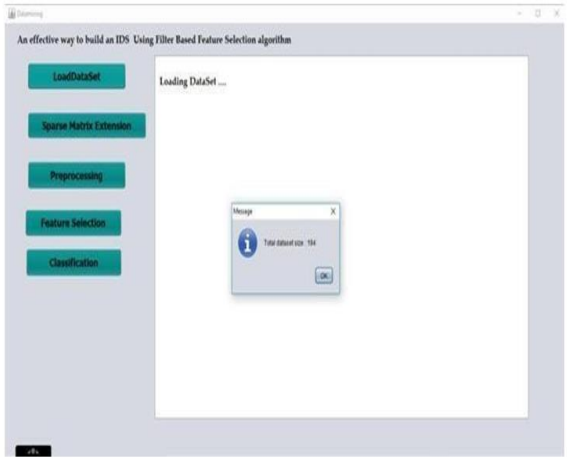


Fig 5.2 KDD cup99 dataset size

2. If we click on OK then KDD cup 99 dataset is loaded and displayed on the screen.

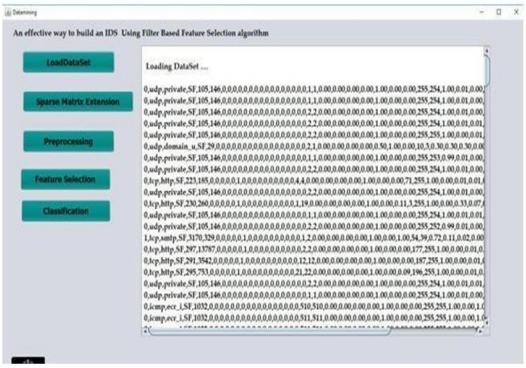


Fig 5.3 Loading KDD cup99 dataset

3. **Sparse Matrix Extension:** Once dataset is loaded,, we reduce the size ofthe dataset by clicking on sparse matrix extension.

Fig 5.4 Size after Sparse matrix

4. **Pre processing:** Two steps are performed under pre processing, data transfer anddata normalization. If we click on DataTransfer button, then every symbolic feature in a dataset is first converted into numerical value

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

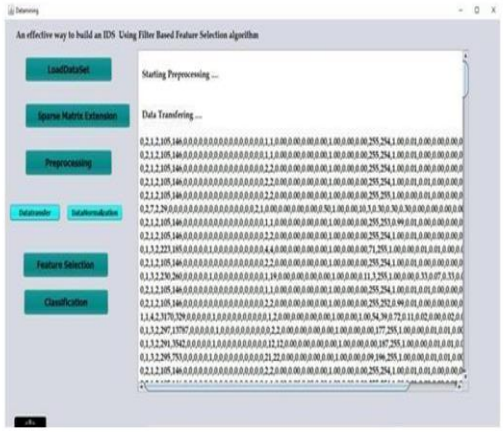


Fig 5.5 Data Transferring

5. **Data Normalization:** After data transferring, data is normalized**.** Data normalization is a process of scaling the values of each record into a well proportioned range, so that the bias in favour of features with greater values is eliminated from the data set.

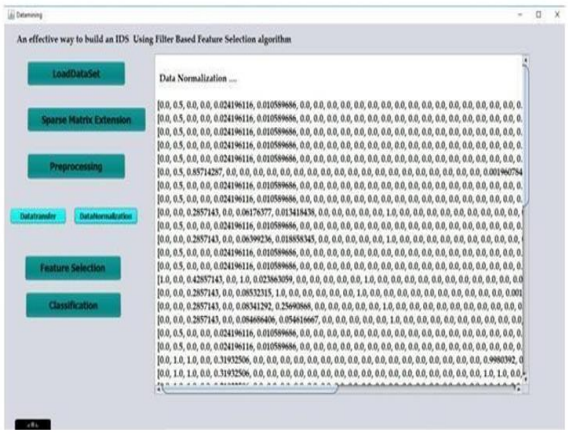


Fig 5.6 Data Normalization

6. **Feature Selection:** Even though every connection in a dataset is represented by various features, not all of these features are needed to build an IDS. Therefore, it is important to identify the most informative features of traffic data to achieve higher performance. In the previous section using Algorithm 1, a flexible method for the problem of feature selection, FMIFS, is developed.

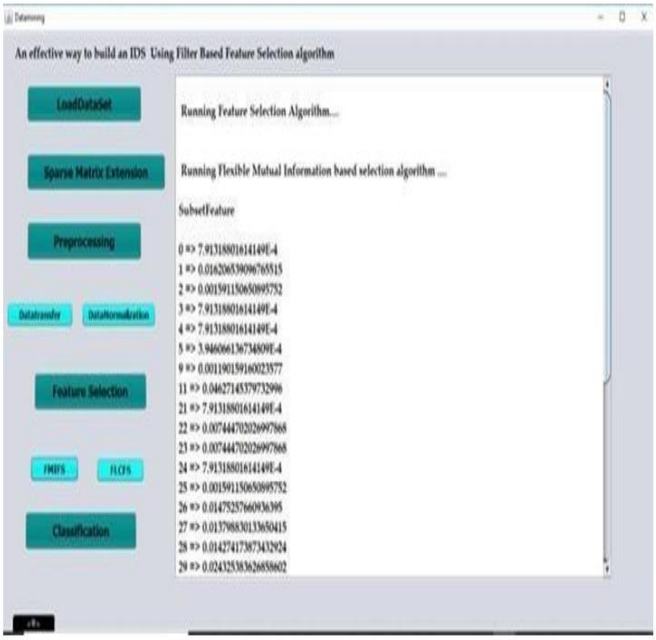


Fig 5.7 Running Flexible Mutual Information based Feature Selection (FMIFS) algorithm

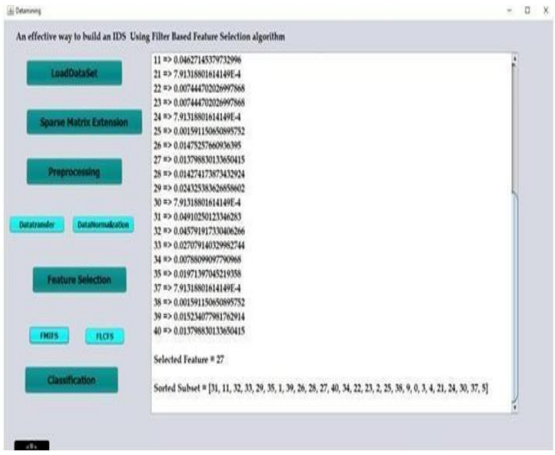


Fig 5.8 Sorted subset of selected features

7. The another feature selection algorithm used over here is flexible linear correlation coefficient based feature selection.

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

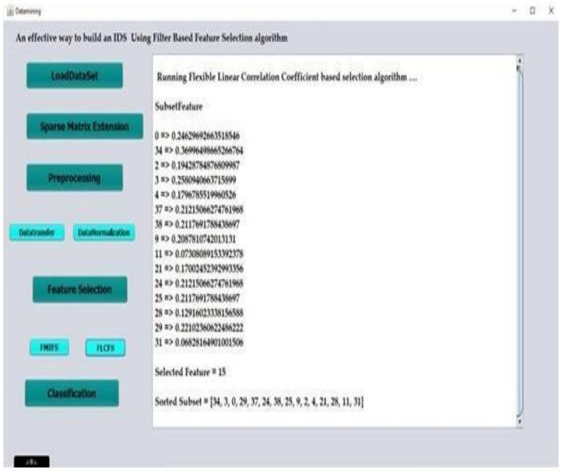


Fig 5.9 Running Flexible Linear Correlation coefficient based (FLCFS) algorithm

8. **Classification:** Once the optimal subset of features is selected, this subset is then taken into the classifier training phase where LS-SVM is employed. Since SVMs can only handle binary classification problems and because for KDD Cup 99 five optimal feature subsets are selected for all classes, five LS-SVM classifiers need to be employed. Each classifier distinguishes one class of records from the others. For example the classifier of Normal class distinguishes Normal data from non-Normal (All types of attacks).

CONCLUSION

I conclude that two main components are essential to build an IDS. They are a robust classification method and an efficient feature selection algorithm. A supervised filter-based feature selection algorithm has been proposed, namely Flexible Mutual Information Feature Selection (FMIFS). FMIFS is an improvement over MIFS and MMIFS. FMIFS suggests a modification to Battiti's algorithm to reduce the redundancy among features. FMIFS eliminates the redundancy parameter β required in MIFS and MMIFS. This is desirable in practice since there is no specific procedure or guideline to select the best value for this parameter.

FMIFS is then combined with the LSSVM method to build an IDS. The proposed LSSVMIDS + FMIFS has been evaluated using well known intrusion detection datasets: KDD Cup 99. The performance of LSSVM-IDS + FMIFS on KDD Cup test data has exhibited better classification performance in terms of classification accuracy, detection rate, false positive rate and F-measure than some of the existing detection approaches.

Finally, based on the experimental results achieved on all datasets, it can be concluded that the proposed detection system has achieved promising performance in detecting intrusions over computer networks.

I further extend this system by implementing sparse matrix concept using which we can remove erroneous and unrelated data. Some time in data mining some source of data generators (sensors sensing temperature or IDS system which will record network logs) will generate error or unrelated data and this data will put effect when performing classification.

This unrelated data we can remove using sparse matrix concept which compare similarity between two records, if two records are completely unrelated then there will be less similarity and it will be remove out.

Using this technique dataset size can be reduced and can improve classification.

FUTURE SCOPE

In further development a more effective data reduction technique can be used. Although the proposed feature selection algorithm FMIFS has shown encouraging performance, it could be further enhanced by optimizing the search strategy. In addition, the impact of the unbalanced sample distribution on an IDS needs to be given a careful consideration in our future studies.

REFERENCES

1. Mohammed A. Ambusaidi, Member, IEEE, Xiangjian He*, Senior Member, IEEE, Priyadarsi Nanda, Senior Member, IEEE, and Zhiyuan Tan, Member, IEEE, ―Building an intrusion detection system using a filter-based feature selection algorithm‖, IEEE TRANSACTIONS ON COMPUTERS, VOL., NO NOVEMBER 2014

2. S. Pontarelli, G. Bianchi, S. Teofili, ―Traffic-aware design of a high speed fpga network intrusion detection system‖, Computers, IEEE Transactions on 62 (11) (2013) 2322–2334.

3. B. Pfahringer, ―Winning the kdd99 classification cup: Bagged boosting‖, SIGKDD Explorations 1 (2) (2000) 65–66.

4. I. Levin, ―Kdd-99 classifier learning contest: Llsoft's results overview‖, SIGKDD explorations 1 (2) (2000) 67–75.

5. D. S. Kim, J. S. Park, ―Network-based intrusion detection with support vector machines‖, in: Information Networking, Vol. 2662, Springer, 2003, pp. 747–756.

6. A. Chandrasekhar, K. Raghuveer, ―An effective technique for intrusion detection usingneuro-fuzzy and radial svm classifier‖,in: Computer Networks & Communications (NetCom), Vol. 131, Springer, 2013, pp. 499–507.

7. S. Mukkamala, A. H. Sung, A. Abraham, ―Intrusion detection using an ensemble of intelligent paradigms‖, Journal of network and computer applications 28 (2) (2005) 167–182.

8. A. N. Toosi, M. Kahani, ―A new approach to intrusion detection based on an evolutionary soft computing model using neuro fuzzy classifiers‖, Computer communications 30 (10) (2007) 2201–2212.

9. Z. Tan, A. Jamdagni, X. He, P. Nanda, L. R. Ping Ren, J. Hu, ―Detection of denial-of-service attacks based on computer vision techniques‖, IEEE Transactions on Computers 64 (9) (2015) 2519–2533.

10. A. M. Ambusaidi, X. He, P. Nanda, ―Unsupervised feature selection method for intrusion detection system‖, in: International Conference on Trust, Security and

11. A. M. Ambusaidi, X. He, Z. Tan, P. Nanda, L. F. Lu, T. U. Nagar, ―A novel feature selection approach for intrusion detection data classification‖, in: International.

12. Ansam Khraisat, Iqbal Gondal, Peter Vamplew and Joarder Kamruzzaman Survey of intrusion detection systems: Techniques, datasets and challenge. Khraisat et al. Cybersecurity (2019).