Feature Selection Algorithm for Constructing Intrusion Detection System

K.V.Ravi Kiran1, Mr. B.v.v.Satyanarayana Rao 2

1PG Scholar, Dept of CSE, MalineniLakshmaiah Engineering College, Singarayakonda, Prakasam(Dt), AP, India.

2Asst Professor, Dept of CSE, MalineniLakshmaiah Engineering College, Singarayakonda, Prakasam(Dt), AP, India.

Abstract_ Abundance and unnecessary features in the large amount of data have caused a problem in data traffic classification which in turn slowdowns the classification process. Not only does this it also not allow the classifier for making extract decisions, which play a major role in big data. This system uses an algorithm based on mutual information which in turn selects the optimal features for classifications analytically, since it can handle linear and non linear features. Its efficiency can be evaluated in the network detection system. An Intrusion Detection System (IDS) named Least Square Support Vector Machine is fabricated using the feature selected by the algorithm. The performance of LSSVM-IDS can be obtained using three kinds of dataset namely KDD Cup 99, NSL-KDD and Kyoto 2006 dataset. The results show that algorithm contributes more critical features for the LSSVM-IDS to accomplish better exactness and lower computational cost.

Keywords: Big data, Classifier, Intrusion Detection System, Performance, Support Vector Machines.

I.INTRODUCTION

As the years have passed by computer attacks have become less glamorous. Just having a computer or local network connected to the internet, heightens the risk of having perpetrators try to break in, installation of malicious tools and programs, and possibly systems that target machines on the internet in an attempt to remotely control them. The (GOA) team categorised the attacks encountered in 2014 discovering that 25% of the attacks where non-cyber threats followed by scan/probes/attempted access 19% and policy violation 17% This data is further acknowledged by the annual FBI/CSI survey which discovered that though virus based attacks occurred more frequently, attacks based on unauthorised access and denial of service attacks both internally as well as externally, increased drastically.

Recent exploits also suggest that the more sensitive the information that is held is, the higher the probability of being a target. Several Retailers, banks, public utilities and organizations have lost millions of customer data to attackers, losing money and damaging their brand image [2]. In some cases attackers steal sensitive information and attempt to blackmail companies by threatening to sell it to third parties

In the second quarter of 2014, Code Spaces (source code company) was forced out of business after attackers deleted its client databases and backups. JP Morgan, Americas‟ largest bank, suffered a cyber-attack in 2014 that impacted 76 million members [3]. In 2014, Benesse, A Japanese Education Company for children suffered a major breach whereby a disgruntled former employee of a third-party partner disclosed up to 28 million customer accounts to advertisers [4]. Most notably the "Sony Pictures hack" best displayed how significant a companies‟ losses are in the aftermath of a security breach. The network servers were temporarily shut down due to the hack [4]. Cybersecurity experts estimate that Sony lost up to $100 million [5] [6]. Other companies under the Sony blanket fell victim to attacks [7]. To tackle this growing trend in computer attacks and respond threat, industry professionals and academics are joining forces in a bid to develop systems that monitor network traffic activity raising alerts for unpermitted activities. These systems are best described as Intrusion Detection Systems.

II. RELATED WORKS

Existing arrangements stay unequipped for completely ensuring web applications and PC systems against the dangers from constantly progressing digital assault methods, for example, DOS assault and PC malware. Current system movement information, which are regularly tremendous in size, show a noteworthy test to IDSs. These "huge information" back off the whole recognition prepare and may prompt unacceptable arrangement precision because of the computational challenges in taking care of such information. Arranging a gigantic measure of information more often than not causes numerous scientific troubles which then prompt higher computational intricacy. Vast scale datasets for the most part contain loud, excess, or uninformative elements which introduce basic difficulties to learning disclosure and information demonstrating. Chandrasekhar, K. Raghuveer et al suggested that Intrusion recognition is not yet a flawless innovation. The chance to make a few imperative commitments to the field of interruption recognition utilizing information mining Concepts [05]. In this framework, framework has proposed another strategy by using information mining methods, for example, neuro-fluffy and spiral premise bolster vector machine (SVM) for the interruption discovery framework. Their proposed method has four noteworthy strides in which, initial step is to play out the Fuzzy Cimplies bunching. At that point, neuro-fluffy is prepared, to such an extent that each of the information point is prepared with the comparing neuro-fluffy classifier related with the bunch. In this manner, a vector for SVM grouping is framed and in the fourth step, order utilizing spiral SVM is performed to recognize interruption has happened or not. Informational index utilized is the KDD glass 99 dataset and framework has utilized affectability, specificity and precision as the assessment measurements parameters. Our system could accomplish better precision for a wide range of interruptions. It accomplished around 98.94 precision in the event of DOS assault and achieved statures of 97.11 exactness if there should arise an occurrence of PROBE assault.

S. Mukkamala, A. H. Sung, A. Abraham et al proposed Soft figuring procedures are progressively being utilized for critical thinking. This framework addresses utilizing an outfit approach of various delicate processing and hard figuring strategies for interruption location. Because of expanding episodes of digital assaults, building successful interruption discovery frameworks are fundamental for ensuring data frameworks security, but then it remains a slippery objective and an awesome test. Framework concentrated the execution of Artificial Neural Networks (ANNs), Support Vector Machines (SVMs) and Multivariate Adaptive Regression Splices (MARS). Framework demonstrates that a group of ANNs, SVMs and MARS is better than individual methodologies for interruption discovery [6].

III. METHODOLOGY

An intrusion detection system (IDS) is a device or software application that monitors a network or systems for malicious activity or policy violations. Any detected activity or violation is typically reported either to an administrator or collected centrally using a security information and event management (SIEM) system.

A. Intrusion Detection Framework on Least Square Vector Machine

The framework of the proposed intrusion detection system is depicted in figure 1. The detection framework is comprised of four phases: (1) data collection (2) data preprocessing (3) classifier training, and (4) attack recognition.

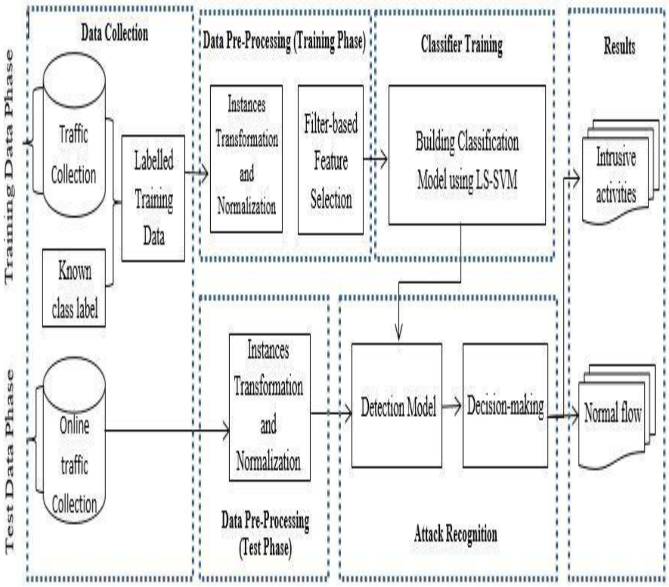
Training Data Phase
Test Data Phase
Data Collection
Traffic Collection
Known class label
Labelled Training Data
Online traffic Collection
Data Pre-Processing (Training Phase)
Instances Transformation and Normalization
Filter-based Feature Selection
Classifier Training
Building Classification Model using LS-SVM
Results
Intrusive activities
Normal flow
Instances Transformation and Normalization
Data Pre-Processing (Test Phase)
Detection Model
Decision-making
Attack Recognition

Fig 1: The framework of the LS-SVM-based Intrusion Detection System.

Data Collection

In data collection we collect a data from the KDD Cup 99 dataset where data is collected based on two factors design and effectiveness of IDS.

Data Preprocessing

In data preprocessing we have stages which are explained below

Data Transferring

Here in data transferring ever symbolic feature in dataset is converted to integer type. For example, the KDD CUP 99 dataset contains both symbolic feature and integer type, these symbolic features include TCP,UDP further it is replaced with integer type.

Data Normalization

There are 3 types of normalization step. In first one we delete all the duplicate data and unwanted data. In second step we generalize the some of the field values. In third step we put zero for the entire field which do not contain any value or which are empty. With help of this step comparison becomes easy.

Feature Selection

In Feature Selection the values which system have got are compared with trained dataset and only some features are selected based on the algorithm flexible mutual information based feature selection and flexible linear correlation coefficient based feature selection.

Attack recognition

In this there are two main steps, in first step the system takes the data and compare with the trained dataset and recognitions if the data is attacked or normal data. If the data is attack then it will undergo the second step, in which the attacked data is classified according to which type of attack is occurred by comparing it with the trained dataset.

V. CONCLUSION

The two main components to build an IDS are robust classification and feature selection .As proposed an algorithm namely Flexible Mutual Information Feature Selection (FMIFS) supervised by Filter Based feature selection algorithm .FMIFS modifies the Battitis algorithm which redundancy among the features and eliminated the redundancy used in MIFS and MMIFS. There is no pre described procedure to select value. FIMS+LSSVM is used to build an IDS. The proposed LSSVM-IDS+FMIFS has been evaluated here with the help of KDDCUP 99 data set .But we get many other datasets like NSL-KDD and Kyoto 2006+datasets for evaluation The corrected set of data of KDD cup 99 data set are tested on normal, DOS and probe classes .The performance is evaluated in the term of accuracy, detection rate, False positive and F-measure. Finally, based on the experimental results achieved on KDD CUP 99.So the result of the system is achieved promising performance in detecting intrusions in the network.

REFERENCES

[1] S. Pontarelli, G. Bianchi, and S. Teofili, “Traffic-aware design of a high-speed FPGA network intrusion detection system,” IEEE Trans. Comput., vol. 62, no. 11, pp. 2322–2334, Nov. 2013.

[2] B. P fahringer, “Winning the KDD99 classification cup: Bagged boosting,”
SIGKDD Explorations Newslett., vol. 1, no. 2, pp. 65–66, 2000.

[3] S. Mukkamala, A. H. Sung, and A. Abraham, “Intrusion detectionusing an ensemble of intelligent paradigms,” J. Netw. Comput.Appl., vol. 28, no. 2, pp. 167–182, 2005.

[4] C. Grosan, C. Martin-Vide, A. Abraham, "Evolutionary Design of Intrusion Detection Programs", International Journal of Network Security, vol. 4, pp. 328-339, 2007.

[5] A. N. Toosi and M. Kahani, “A new approach to intrusion detectionbased on an evolutionary soft computing model using neurofuzzyclassifiers,” Comput. Commun., vol. 30, no. 10, pp. 2201–2212, 2007.

[6] Z. Tan, A. Jamdagni, X. He, P. Nanda, L. R. Ping Ren, and J. Hu,“Detection of denial-of-service attacks based on computer visiontechniques,” IEEE Trans. Comput., vol. 64, no. 9, pp. 2519–2533,Sep. 2015.

[7] A. M. Ambusaidi, X. He, and P. Nanda, “Unsupervised featureselection method for intrusion detection system,” in Proc. Int.Conf. Trust, Security Privacy Comput Commun., 2015, pp. 295–301.

[8] A. M. Ambusaidi, X. He, Z. Tan, P. Nanda, L. F. Lu, and T. U.Nagar, “A novel feature selection approach for intrusion detectiondata classification,” in Proc. Int. Conf. Trust, Security Privacy Comput.Commun., 2014, pp. 82–89.

Author’s Profile

K.V.Ravi Kiran completed M.C.A in 2007 Horizon institute of technology in Hyderabad ,Rangareddy (Dist) and pursing M.Tech in Malineni Lakshmaiah

Engineering college ,Singarayakonda,
Prakasam (Dist).AP, India.

Mr.P.Srinivasulu Has Received His B.Tech In CSE In And M.Tech PG In Computer Science. He Is Dedicated To teaching Field From The Last 7 Years. At Present He Is Working As Assistant Professor In Malineni Lakshmaiah Engineering College, Singarayakonda, Prakasam(Dt), AP, India.. He Is Highly Passionate And Enthusiastic About Her Teaching And Believes That Inspiring Students To Give Of His Best In Order To Discover What He Already Knows Is Better Than Simply Teaching.