Real Time Anomaly Intrusion detection using SVM and Genetic selection process and Fuzzy Logic

Mr. Jitendra Soni
Asst.Professor
Dept.Of Computer Science
IET-DAVV
INDORE, INDIA

Mr. Arpit Agrawal
Asst.Professor
Dept.Of Computer Science
IET-DAVV
INDORE, INDIA

Govind Bisen
M.E Final Year
Dept.Of Information Security
IET-DAVV
INDORE, INDIA

***Abstract*: There are many approaches for Intrusion Detection, HIDS and NIDS are two prime approaches and all other approaches are subpart of HIDS and NIDS. Host based Intrusion Detection System and Network based Intrusion Detection System have many approaches, Stronger security methods like security methods such as advanced encryption algorithms, efficient authentication process, System call based Intrusion Detection are some of the Efficient approaches. Despite all these approaches cyber security need to be upgraded because Anomaly based Intrusion are undetected, they are dynamic in nature. So the approaches to detect these attacks should be sequential as well as random. Haste makes waste so fast and inefficient algorithms have no use. It should be secure first and second it should be fast. In this paper we will use fuzzy technique with genetic selection process to reduce data and SVM (Support Vector Machines). SVM will classify the data to identify Intrusion.**

Key Words: SVM, HIDS, NIDS, INTRUSION, REDUCED GENETIC SELECTION

1 Introduction:

At Present almost all the transaction happens through Internet, Hacker try to use transaction in the Internet that may steal the information. It is very difficult to separate the original useful data with malicious data over the Internet.

Machine learning [1], Data mining, and soft computing are some of the most widely used technique used to identify the intruder's work. Intrusion detection System [2] is an Artistic work to identify Intruder. Alarm rate is one of the important factor depends on security of IDS, and its efficiency. Biggest challenge is to reduce False Alarm rate, not to compromise with the security but to increase the efficiency of the IDS. In this paper we will try to improve efficiency of the existing IDS without compromising to security as well as decreasing false alarm rate. This paper introduced to give a concise description to a hybrid feature selection technique for intrusion detection in real time environment.

2 Previous works:

Types of IDS in short

Active intrusion detection system, also known as active intrusion detection and prevention system. It works on run time and identifies the abnormal activity of intruder.

Passive intrusion detection system only monitors the network and analyze, it is not capable of performing any protective or potential vulnerability and attacks but the major advantage of passive IDS they are it is not vulnerable to attacks.

Not getting into depth but there are some other intrusion detection system called Host based IDS, Network based IDS, and Knowledge based IDS and behavior based IDS. Log based intrusion detection system based on the database or logs of previous attacks, this database need to be updated regularly, but it may not stop new type of attacks, for these new type of attacks there is anomaly based intrusion detection system that also works on the behavior of the system and generate alarm on changing behavior of the system, this may generate false alarm rate.

3 Related works

In this section briefly outline of the concepts used in this paper

KDD CUP 99 Dataset

The 1998 DARPA Intrusion Detection Evaluation Program was prepared and managed by MIT to evaluate research in Intrusion detection technology

ISSN: 2278 – 1323

[6][9], it is a standard set of data not quite big, not too small, and it is multidimensional comprises of multiple attacks on military environment .

Weeks of raw TCP dump data for a local-area network (LAN) simulating a typical U.S. Air Force LAN by Lincoln lab. Real time environment was simulated.

A connection is a sequence of TCP packets starting and ending at some well-defined times, between which data flows to and from a source IP address to a target IP address under some well-defined protocol. Each connection is labeled as either normal, or as an attack, with exactly one specific attack type. Each connection record consists of about 100 bytes.

There are four main type of attacks:

1) DOS: denial-of-service, e.g. syn flood;
2) R2L: unauthorized access from a remote machine, e.g. guessing password;
3) U2R: unauthorized access to local super user (root) privileges, various buffer overflow" attacks;
4) Probing: surveillance and other probing, e.g., port scanning.

There are 42 columns with 41 attributes and one attribute shows if it is attack or normal.

Preprocessing:
Data that has to be analyze contains redundant information and useless data, in order to make it to ready for experiment, it is required to remove those redundant and useless information, this process contains KDD CUP99 dataset we can use discretization technique to process the data, SVM[8] can't handle such large chunk of information, it will then reduce the data easy for mapping over SVM.

Feature selection:
Large dataset may contain vast pool of Information, some feature may be useless or may be predictive, and to reduce the dimensionality of data by selecting only a subset of measured feature to create a model is called Feature Selection. This is a hybrid approach of feature selection.

Several techniques are available to reduce the feature
Negative Matrix Factorization
Principal Component Analysis
Factor Analysis
Randomized feature selection
Sequential feature selection etc.

Fuzzy logic:
Fuzzy logic is a logical system which is an extension of multivalued logic. It is a more intuitive approach without the far-reaching complexity. It is also related to fuzzy sets, that relate the classes of objects with boundaries and membership of the object in particular class is a matter of degree. In our proposed method we are going to use Fuzzy logic for selecting best feature from the set of features.

SVM and kernel trick:
There is a class of functions G(x,y) with the following property. There is a linear space S and a function φ mapping x to S such that

$G(x,y) = \langle \varphi(x), \varphi(y) \rangle$.

The dot product takes place in the space S.

Classification problems don't have a simple hyper plane as a useful separating criterion. For those problems, there is a variant of the mathematical approach that retains nearly all the simplicity of an SVM separating hyper plane.

Polynomials: For some positive integer d
$G(x,y) = (1 + x'y)d$.

Radial basis function (Gaussian): For some positive number σ,

$G(x,y) = \exp(-(x-y)'(x-y)/(2\sigma 2))$.

Multilayer perceptron (neural network): For a positive number p1 and a negative number p2,

$G(x,y) = \tanh(p1x'y + p2)$.

Test result will be obtained by
FAR:
False alarm rate total no of normal process / total process
Accuracy rate:
Alarm rate for Intruder's process / total process

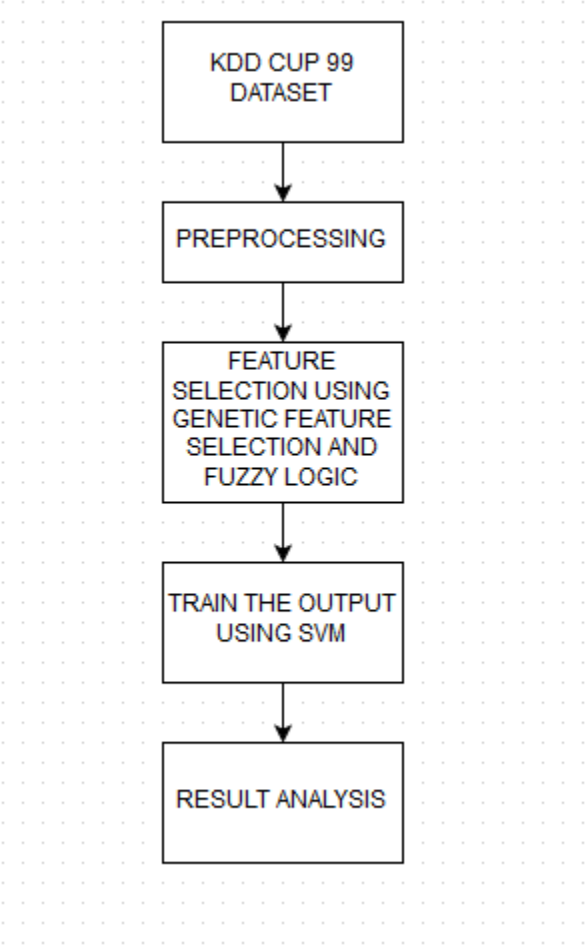
DESIGN AND IMPLEMENTATION:

Creating an anomaly detection system is challenging task as the algorithm should be very efficient some optimization algorithm are really very efficient but because accuracy is not so good, it may cause false alarm rate, we are trying to make an efficient method to reduce the dataset dimension without making it less efficient and vulnerable.
Method is based on following steps;

1) KDD CUP99 Dataset as input.
2) Preprocessing of the data

ISSN: 2278 – 1323

3) Genetic algorithm and fuzzy logic based feature selection
4) Application of SVM
5) Result analysi



Feature selection	No of feature	Accuracy
Proposed method	16(expected)	95%(expected)
PCA	12	80%

Conclusion:

In this work, the Intrusion Detection approach uses genetic and fuzzy logic based selection and the SVM kernel trick. Feature selection reduces the dimensionality of the input space, thus increases the system performance and decreases the memory usage, it saves lot of computation resources. It will also reduce the feature from 41 to 16 and will save precious time for identify Intrusion. Hybrid selection process will provide accuracy with efficiency.

REFERENCES:

[1] Lee W and Stolfo S., "Data Mining techniques for intrusion detection", In: Proc. of the 7th USENIX security symposium, San Antonio, TX, 1998.

[2] Denning D. (1987) "An Intrusion-Detection Model," IEEE Transactions on Software Engineering, Vol. SE-13, No. 2, pp.222-232.

[3] LIBSVM -- A Library for Support Vector Machines:www.csie.ntu.edu.tw/~cjlin/libsvm/

[4] http://svms.org/kernels/

[5] en.wikipedia.org

[6]**http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html**

[7] Boussouf M (1998) A Hybrid Approach to Feature Selection. Lecture Notes in Artificial Intelligence 1510:231–238.

[8] Xu P and Chan A., An efficient algorithm on multi-class support vector machine model selection. Proceedings of the International Joint Conference on Neural Networks,

[9] Tavallaee M, Bagheri E, Wei Lu, Ghorbani A., "A detailed analysis of the KDD CUP 99 data set," Computational Intelligence for Security and Defence Applications, 2009. CISDA 2009. IEEE Symposium on , vol., no., pp.1,6, 8-10 July 2009

ISSN: 2278 – 1323