Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X

IJCSMC, Vol. 2, Issue. 6, June 2013, pg.110 – 115

RESEARCH ARTICLE

A Novel Fuzzy Logic Analysis & Study on Intrusion Detection System

G. Shilpa[1], N. Anjaneyulu[2]
[1]Assistant Professor, Sri Indu College of Engineering and Technology, Hyderabad, India
[2]Assistant Professor, Sphoorthy Engineering College, Hyderabad, India

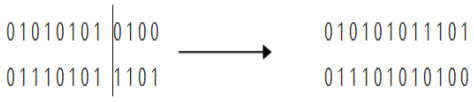
Abstract— Classification of intrusion attacks and normal network activity is increasing problem in computer network security. In this paper, we present a novel intrusion detection approach to extract both accurate and interpretable fuzzy classifier from computer network data. The proposed fuzzy rule-based system is evolved from a feature selection KDD Cup framework. In addition, the proposed system presents the genetic feature selection wrapper to search for an optimal feature subset for dimensionality reduction. To evaluate the intrusion detection classification and feature selection performance of our approach, it is compared with some well-known data mining classifiers as well as feature selection. The comparative results on the KDD-Cup99 intrusion detection benchmark data set demonstrate that the proposed approach produces interpretable fuzzy systems, and outperforms other classifiers by providing the highest detection accuracy for intrusion attacks and low false alarm rate for normal network traffic with minimized number of features.

Key Terms: - Fuzzy classifier; Genetic algorithms; Feature selection; Intrusion detection

I. INTRODUCTION

Mining technologies provides the basic knowledge of fuzzy systems neural networks and genetic algorithms. Genetic algorithms are adaptive procedures derived from the principal survival of the fittest in natural genetics and maintains a potential solutions of the candidate termed as individuals, starts with randomly created initial population of individuals involves encoding of every variable. Binary variables are mapped to real numbers in numerical problems. Genetic algorithm used to solve many combinatorial optimization such as 0/1 knapsack problem, travelling salesperson problem scheduling problems, etc. Binary encoding is suitable to solve many [1] of these problems other than utilized. Selection method in genetic selects parents from the population on the basis of fitness of individuals. High fitness individuals are selected with higher probability of selection to reproduce offsprings for the next population.

In a standard genetic algorithm two parents are selected at a time and are used to create two new children to take into next generation, the offsprings are subject to crossover operator with a pre-specified probability of crossover. It marks a random crossover spot within the size of chromosome and exchanges the bits on the right of the spot as shown here.



Mutation operator is applied to the entire child after crossover. It flips each bit in the individual with a pre-specified probability of mutation.

$$011101010111 \longrightarrow 011111010111$$

Procedure is repeated till number of individuals in the population is complete, finishes one generation in genetic algorithm. Standard genetic algorithm utilizes operators such as reproduction crossover and mutation. The values of genetic parameters such as population size crossover probability mutation probability [2] total number of generations affect convergence properties of the genetic algorithms.

II. RELATED WORK

Mining in data used to analyze complex datasets and useful classification patterns in the datasets. Research like agricultural biometrics studies have used various techniques of data mining including natural trees, statistical machine learning techniques. Genetic algorithm has been widely used in data mining. Effective genetic fuzzy classification, fuzzy clustering are analyzed on the collected as a machine learning soil data which deals with the categorization of soils based on distinguishing characteristics as well as criteria that dictate choices in use. Shah A Kusiak 2004 have applied genetic algorithm for features selection for mining association rules genomic studies provide large volumes of data with thousands of single nucleotide polymorphisms. A fuzzy classification rule is a fuzzy if then rule whose consequent is a class lable since the comprehensibility of rules by human understandable is criterion in designing a fuzzy rule based system, fuzzy classification rules with linguistic interpretations. To detect with this problem we consider the genetic fuzzy mining techniques are used to describe each sample object in dataset. Fuzzy sets are basic concepts supporting fuzzy theory such as fuzzy logic measure etc. although most applications of fuzzy theory have been biased toward engineering applications recently disciplines such as medical diagnostics psychology education economy and a synthetic concept of emotion impression intuition and other human subjective factors.

III. GENETIC ALGORITHM NETWORK INTRUSION DETECTION

3.1 Genetic Algorithm Network Intrusion Detection: The process of GA usually begins with random selection of chromosomes are representations of the problem to be solved. According to the attributes of the problem different positions of each chromosome are encoded as bits, characters or numbers.

Genetic algorithm to intrusion detection is a promising issues, GA can be used to evolve rules for network traffic these rules are used to differentiate normal network connections from anomalous connections refer to events with probability of intrusion.

if {Condition} then {act}

this refer to a match between current network connection and the rules in intrusion detection such as source and destination IP addresses and port numbers used in are TCP/IP protocols duration of the connection protocol indicating the probability of an intrusion. The act field usually refers action defined by the security policies within an organization such as reporting an alert to the system administrator stopping connection, logging a message into system audit files.

For the above condition we require source IP address, destination IP address, port number connection time and stop the connection. The final goal of genetic algorithm conditions that match only to network intrusion connections.

Each parameters influences the effectiveness of the genetic algorithm, the evaluation function is one of the most important parameters. The overall outcome is calculated based on whether a field of the connection matches the object in dataset and then multiply the weight of that field. Matched value is either 0 or 1.

$$\text{Outcome} = \sum \text{Matched} * \text{Weight}$$

The order of weight values categorized according to different fields in the connection record by network traffic therefore all objects in representing destination IP address field have same weight. Destination IP address is the target of an intrusion while the source IP address is the originator of the intrusion.

3.2. Design of Fuzzy Logic Control: Fuzzy controllers compared to the classical controllers such as the simplicity of control low cost and the possibility to design without knowing the exact mathematical model of the process. Fuzzy logic is an application of fuzzy set in which the variables are linguistic rather than the numeric variables. Linguistic variables defined as variables whose values are sentences in a natural language may be represented by fuzzy sets. Fuzzy set means element may partially belong to more than one set. A fuzzy set A of a universe of discourse X is represented by a collection of ordered pairs of generic element x€ X and its membership function μ : X ->[0,1] which associates a number μ A(x) : X->[0,1] to each element x of X.

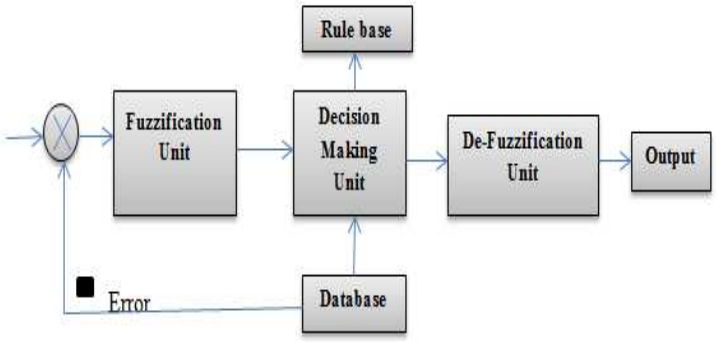


Figure 1 is Fuzzy Classifier

Fuzzification is the first block inside the controller is fuzzification which converts each piece of input data to degrees of membership by a lookup in one or several membership functions, matches the input data with the conditions of the rules to determine how well the condition of each rule matches that particular input instance. Rule base may use several variables both in the condition and the conclusion of the rules be applied to both multi-input-multi-output problems and single-input-single-output problems. Typical SISO problem is to regulate a control signal based on an error signal, may actually need both the error, the change in error and the accumulated error as inputs but we will call it single-loop control because in principle all three are formed from the error measurement. Defuzzification must be converted to a number that can be sent to the process as a control signal type of operation is called defuzzification and the x-coordinates marked by a white vertical dividing line becomes the control signal resulting into several crisp defuzzification methods.

IV. INTRUSION DETECTION

4.1 Intrusion Detection: Intrusion detection is the process of monitoring the events occurring in a computer system. It is defined as attempts to comprise the confidentiality, integrity bypass the security mechanisms of a computer network, intrusion caused by attackers accessing the system from the internet. Intrusion detection allows protecting their system from the threats that come with increasing network connectivity on information systems. It is important to know which intrusion detection features and capabilities to use, to prevent problem behaviours by increasing the perceived risk of discovery and punishment for those who would attack or otherwise abuse the system and to detect attacks and other security violations that are not prevented by other security measures.

There are several types of intrusion detection available to characterize by different monitoring analysis approaches, anomaly detection and misuse detection can be summarized as follows.

Anomaly Detection	**Misuse Detection**
a. Describes normal behavior, and flags deviations	a. Uses a knowledge base to recognize the attacks
b. Uses statistical or machine learning models of behavior	b. Can recognize only attacks for which a "signature" exists in the KB
c. Theoretically able to recognize any attack, also 0-days	c. When new types of attacks are created, the language used to express the rules may not be expressive enough
d. Strongly dependent on the model, the metrics and the thresholds	d. Problems for polymorphism
e. Generates statistical alerts: "Something's wrong"	e. The alerts are precise: they recognize a specific attack, giving out many useful information

Table 1 differentiate the intrusion detection system

There are variable number of intrusion detection system when analysing misuse detection and anomaly detection system, the major approaches of machine learning used in the intrusion detection can be used as fuzzy if then else rule.

4.2 Feature Selection: The feature selection is component of classification tasks recognised by Kittler and Young then existing feature selection procedure which is a plug-in for the commercial software on soft computing data analysis data engine as its roots in fuzzy logic. As the name suggests the significant features and acts as a pre-processor for the classification problems with very high features dimensions. For instance the vibration signals in the time and frequency domain generates a large number of features and makes the reduction of the dimensionality.

The problem of feature selection lies in selecting the best subset Y of n features i.e Y= {yi/i=1,2,…..n} from the total set X={xi/i=1,2,….n} where n<n' classification system if frequently accomplished by separately appraising the feature selection step and inherent classification . Feature selector combines the two steps because it has the advantage that selected features are well suited for the given classifier.

V. PROBLEM DEFINITION

Computer networks play vital role in the society, installation of antivirus software, protect from firewalls the system become more complex, to avoid all these issues our work introduces a system to detect intrusion anomaly detection. The futuristic approach of fuzzy classifier is data driven method forunderlyingdata. Fuzzy classifier guarantees at least a minimum of the criterion function there by accelerating the convergence of rules on KDD Cup datasets.

There are variable numbers of intrusion detection systems where all of those intrusion detections systems can be placed into two major categories when analysing miuse detection and anomaly detection systems. The major approach of fuzzy classifier as an intrusion detection can only detect known attacks, it cannot detect insider attacks or otherwise called as privilege attacks. They do not have holistic picture of the network to detect multi-step attacks over a long time period. Though data for detection is available system administration are limited the better solution for an intrusion system can be fuzzy if then rule which is the process of extracting useful and previously unnoticed models or patterns from large data stores. To be more specific things that might contribute to an intrusion detection using fuzzy classifier.

Remove normal activity from alarm data to allow analysts to focus on real attacks.

Identify false alarm generators and bad signatures and also finds anomalous activity that uncovers a real attack.

Identify long on going patterns different IP address same activity.

To accomplish these tasks we choose one of technique

Data summarization with statistics including findings outliers

Clustering of the data into natural categories

Association rule discovery defining normal activity and enabling the discovery of anomalies
Classification predicting the category to which particular record belongs

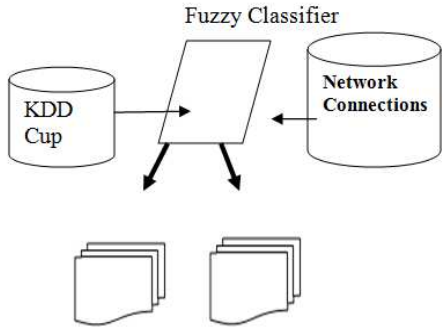


Figure 2 represents the fuzzy classifier learning to detect intrusion anomaly

VI. COMPARATIVE STUDY

The use of neural networks in the area of intrusion detection system has significantly increased. In this we present obtained by comparing the growing fuzzy rule classifier applied to intrusion detection system compared two important aspects the performance and the training time. The results show that the increasing network improves the performance of the system in detection of anomalies obtaining better results between the detection rate and the number of false positive rate, the networks have been trained and tested with data provided by the KDD Cup 99 intrusion detection evaluation.

The fuzzy logic is a problem solving control system that lends to implement in many system like simple small embedded micro-controlled to large networked work stations and can also implemented in hardware, software or a combination of both. Fuzzy classifier requires some numerical parameters in order to operate significant error but exact values of these numbers are usually not critical. Classification in data mining is only used for business analysis but fuzzy logic is applied for software as well as hardware, existing intrusion detection system solved by using machine learning techniques to detect false positive rate, our proposed fuzzy intrusion detection system presents best performance compare to previous work.

VII. CONCLUSION

In this paper, we presents fuzzy classifier to detect intrusion anomaly in human understandable in secure manner, KDD Cup 99 is data set with unauthorized and password authentication. We evaluate the normal or anomalies data classification fuzzy classifier and comparative study is class label comparison of data mining and fuzzy classifier. Fuzzy logic is applicable for both software and hardware. Our work shows best analysis compare existing work. Future analysis extends to investigate on implementation different attacks of intrusion anomalies.

REFERENCES

[1] P. Bhargavi S. Jyothi 2011 soil classification using data mining techniques: a comparative study. International Journal of Engineering Trends and Technology, July- Augest Issue 2011.
[2] Bezroukov, Nikolai. 19 July 2003. “Intrusion Detection (general issues).” Softpanorama: Open Source Software Educational Society. Nikolai Bezroukov. URL: http://www.softpanorama.org/Security/intrusion_detection.shtml (30 Oct. 2003).
[3] Bridges, Susan, and Rayford B. Vaughn. 2000. “Intrusion Detection Via Fuzzy Data Mining.” In Proceedings of 12th Annual Canadian Information Technology Security Symposium, pp. 109-122. Ottawa, Canada.
[4] Crosbie, Mark, and Gene Spafford. 1995. “Applying Genetic Programming to Intrusion Detection.” In Proceedings of 1995 AAAI Fall Symposium on Genetic Programming, pp. 1-8.
[5] Cambridge, Massachusetts. URL: http://citeseer.nj.nec.com/crosbie95applying.html (30 Oct. 2003).
Graham, Robert. Mar. 21, 2000. “FAQ: Network Intrusion Detection Systems.” RobertGraham.com Homepage.
[6] Robert Graham. URL:

http://www.robertgraham.com/pubs/network-intrusion-detection.html (30 Oct. 2003).

[7] Jones, Anita. K. and Robert. S. Sielken. 2000. "Computer System Intrusion Detection: A Survey." Technical Report.Department of Computer Science, University of Virginia, Charlottesville, Virginia.

Authors Bibliography

G.SHILPA received the B.Tech (CSE) from Sri Venkateshwara Engineering College, and the M.Tech (SE) from St.Marry's Engineering College, Hyderabad, India in 2012. She is currently working as Assistant Professor in Sri Indu College of Engineering and Tech, Hyderabad, India. Her area of interests is Computer Networks, Web Applications, Mobile Computing, Data Base Management System, C programming and Data Structures.

N.ANJANEYULU received the B.Tech (IT) from CVR College of Engineering, Hyderabad, and M.Tech (IT) from Guru Nanak Engineering College, Hyderabad, India in 2011. He is currently working as Assistant Professor in Sphoorthy Engineering College, Hyderabad, India. His area of interest includes Computer Networks, Network Security, Mobile Computing and Web Applications.

