**IDS FOR HYBRID CLOUD**

**MUSHTAQ-AHMAD-DAR   
DEPARTMENT-OF-I.T  
MTECH-I.T54-CHANDIANA GANDERBAL**

mushtaqdar@hotmail.com

1. **Abstract:-**

Security of valuable information is always a very essential issue for modern digital world. Intrusion Detection System (IDS) and many security techniques are widely used against cyber-attacks. Data mining and machine learning methods have been used by researchers to obtain high detection rate and low false alarm rate. Proposed work aims to design and development of an approach for improve cyber-attack detection system using cloud.

**1.1 Introduction**:

An Intrusion detection system examines all internal and external network activities or attacks and identifies suspicious design that may point out the system attack or a network from someone attempting to break into the security or compromise a system . Due to the distributed nature of cloud computing, cloud computing environments are easy targets for invaders looking for possible susceptibility to exploit.

**2.1.1 Algorithm Used:**

The most common Shallow Learning

algorithms used for IDS are Decision Tree , K Nearest Neighbor (KNN), Artificial Neural Network (ANN), Support Vector Machine (SVM), K Mean Clustering, Fast Learning Network and Ensemble Methods

**2.1.2 Tools used:**

Snort 3

Suricata

Implementing PGPA

**2.1.3 Execution of Implementation work:**

The experimental steps that will include for the Detection system:-

1. Create New Resource: Machine Learning Analytics solution.

2. Import/Upload the dataset.

3. Pre-process the dataset. Data pre-processing can also be done using modules written ML.

4. Randomly split and partition the data into 70% training and 30% testing, using the ‘Split Data’ module.

5. Identify categorical attributes and cast them into categorical features using the ‘Edit Metadata’ module.

6. Convert to Indicator Values module to convert columns that contain categorical values which can more easily be used as features.

7. Select Columns in Dataset those are relevant

8. Apply Ensemble Method

9. Apply Machine Learning Algorithm to Train the model.

10. Now Score and Evaluate the Model. The ‘Evaluate model’ also visualizes the results through confusion matrix.

**2.2.1 Data Set-**

The Dataset To implement the algorithm and to evaluate the performance of the system, I propose the standard datasets employed in KDD Cup 1999 “Computer Network Intrusion Detection” competition. The KDD 99 intrusion detection datasets depends on the 1998 DARPA proposal, which offers designers of intrusion detection systems (IDS) with a standard on which to evaluate different methodologies ([21], [24]). Hence, a simulation is being prepared from a fabricated military network with three ‘target’ machines running various services and operating systems. They also applied three extra machines to spoof different IP addresses for generating network traffic. A connection is a series of TCP packets beginning and ending at some well-defined periods, between which data floods from a source IP address to a target IP address under some well-defined protocol ([21], [22], [24]). It results in 41 features for each connection. Finally, a sniffer accounts all network traffic by means of the TCP dump format [24]. The total simulated period is seven weeks. Normal connections are shaped to outline that expected in a military network and attacks are categorized into one of four types: User to Root; Remote to Local; Denial of Service; and Probe.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| DATASET | NORMAL | PROB | DOS | U2R | R2L | TOTAL |
| TRAIN | 97280 | 4107 | 391458 | 52 | 1124 | 494021 |
| TEST(“Corrected”) | 60593 | 4166 | 229853 | 228 | 16189 | 311029 |

**2.3. Genetic algorithm**

A Genetic Algorithm (GA) is a programming technique that uses biological evolution as a problem solving strategy [20]. It is based on Darwinian’s theory of evolution and survival of fittest to make effective a population of candidate result near a predefined fitness [13]. The proposed GA based intrusion detection system holds two modules where each acts in a dissimilar stage. In the training stage, a set of classification rules are produced from network audit data using the GA in an offline background. In the intrusion detection phase, the generated rules are employed to classify incoming network connections in the real-time environment. Once the rules are generated, the intrusion detection system becomes simple, experienced and efficient one. The process generally starts with randomly generated population of chromosomes, which signify all possible solution of a problem that are measured candidate solutions. From each chromosome, different positions are set as bits, characters or numbers. These positions are regarded as genes. An evaluation function is employed to find the decency of each chromosome according to the required solution; this function is known as “Fitness Function”. During the process of evaluation, “Crossover” is applied to have natural reproduction and “Mutation” is applied to mutation of species [13]. For survival and combination, the selection of chromosomes is partial towards the fittest chromosomes. When I use GA for solving various problems three factors will have crucial impact on the use of the algorithm and of the applications [2].

The factors are: i) the fitness function, ii) the representation of individuals, and iii) the genetic algorithm parameters. The determination of these factors often depends on implementation of the system. 2.2.2 Fuzzy logic Zadeh explained that Fuzzy logic [9] is an extension of Boolean logic that is often used for computer-based complex decision making. While in classical Boolean logic an element can be either a full member or non-member of a Boolean (sometimes called ”crisp”) set, the membership of an element to a fuzzy set can be any value within the interval [0, 1], allowing also partial membership of an element in a set. A fuzzy expert system consists of three different types of entities: fuzzy sets, fuzzy variables and fuzzy rules. The membership of a fuzzy variable in a fuzzy set is determined by a function that produces values within the interval [0, 1]. These functions are called membership functions. Fuzzy variables are divided into two groups: antecedent variables, that are assigned with the input data of the fuzzy expert system and consequent variables, that are assigned with the results computed by the system. The fuzzy rules determine the link between the antecedent and the consequent fuzzy variables, and are often defined using natural language linguistic terms. For instance, a fuzzy rule can be “if the temperature is cold and the wind is strong then wear warm clothes”, where temperature and wind are antecedent fuzzy variables, wear is a consequent fuzzy variable and cold, strong and warm clothes are fuzzy sets.

The process of a fuzzy system has three steps. These steps are Fuzzification, Rule Evaluation, and Defuzzification. In the Fuzzification step, the input crisp values are transformed into degrees of membership in the fuzzy sets. The degree of membership of each crisp value in each fuzzy set is determined by plugging the value into the membership function associated with the fuzzy set. In the rule evaluation step, each fuzzy rule is assigned with a strength value. The strength is determined by the degrees of memberships of the crisp input values in the fuzzy sets of antecedent part of the fuzzy rule. The defuzzification stage transposes the fuzzy outputs into crisp values. It has been revealed by Baruah [6] that a fuzzy number [a, b, c] can be explained with reference to a membership function μ(x) remaining between 0 and 1, a ≤ x ≤ c. Further, he has extended this definition in the following way. Let μ1(x) and μ2(x) be two functions, 0 ≤ μ2(x) ≤ μ1(x) ≤ 1.He has concluded μ1(x) the fuzzy membership function, and μ2(x) a reference function, such that (μ1(x) – μ2(x)) is the fuzzy membership value for any x. Finally he has characterized such fuzzy number by {x, μ1(x), μ2(x); x ∈ Ω}.International Journal of Distributed and Parallel Systems (IJDPS) Vol.4, No.2, March 2013 41

i=1

q

i=1

p

i=1

p

i=1

q

The complement of μx is always counted from the ground level in Zadehian’s theory [9], whereas it actually counted from the level if it is not as zero that is the surface value is not always zero. If other than zero, the problem arises and then we have to count the membership value from the surface for the complement of μx. Thus I could conclude the following statement –Complement of μx = 1 for the entire level Membership value for the complement of μx = 1- μx

In the two classes’ classification problem, two classes are available where every object should be classified. These classes are called positive (abnormal) and negative (normal). The data set employed by the learning algorithms holds a set of objects where each object contains n+1 attributes. The first n attributes identifies the monitored parameters of the object characteristics and the last attribute identifies the class where the object belongs to the classification attribute. A fuzzy classifier is a set of two rules for solving the two classes’ classification problem, one for the normal class and other for the abnormal class, where the conditional part is described by means of only the monitored parameters and the conclusion part is viewed as an atomic expression for the classification attribute.

**2.4.1 Detection Accuracy**

The accuracy of intrusion detection system relies on the technique in which it identifies, such as by the rule set. Signature-based detection detects only simple and recognized attacks, while anomaly-based detection can detect more types of attacks, but has a higher number of false positives ratios. Tuning is essential to reduce the number of false positives and to make the data further functional.

**2.4.2 Signature database** - A common policy for IDS in detecting intrusions is to remember signatures of known attacks. The inherent weak points in relying on signatures are that the signature patterns must be acknowledged first. New threats are often unrecognizable by eminent and popular IDS. Signatures can be masked as well. The ongoing event between new attacks and detection systems has been a challenge. Therefore the signature database must be updated whenever a different kind of attack is detected and repair for the same is available.

Monitor traffic in large networks - Network Intrusion Detection System (NIDS) components are spotted throughout a network, but if not placed tactically, many attacks can altogether avoid NIDS sensors by passing through alternate ways in a network. Moreover, though many IDS products available in the market are efficient to distinguish different types of attacks, they may fail to recognize attacks that use many attack sources. Many IDS cannot cleverly correlate data from numerous sources. Newer IDS technologies must influence integrated systems to increase an overview of distributed intrusive activity. Therefore IDS must be able to successfully monitor traffic in a large network.

2.4.5 **CONCLUSION**

I have described an overview of some of the current and past intrusion detection technologies, which are being utilized for the detection of intrusive activities against computer systems or networks. The different detection challenges that affect the decision policy of the IDS employed in an organization are clearly outlined. I propose to use the new definition of the complement of fuzzy sets where the fuzzy membership value and fuzzy membership function for the complement of a fuzzy set are two different concepts because the surface value is not always counted from the ground level. By joining much more principle the effectiveness of the method can be enhanced to discover an Intrusion in a network By improving the values based upon the limitation of time the implementation of the method can be enhanced If the counting of users gets enlarged, the performance of the proposed method will hold well

**3.References:-**

1. M.M.M.Hassan Current Studies on Intrusion Detection System, Genetic Algorithm and Fuzzy Logic”, International

Journal of Distributed and Parallel Systems, Vol.4, No.2, pp.35 47

1. Ms. Parag K Shelke , Ms. Sneha Sontakke and Dr. A. D. Gawande , “Intrusion Detection System for Cloud Computing”,

International Journal of Scientific & Technology Research,Vol.1,ISSN 2277 8616, 2012.

1. Hassen Mohammed Alsafi , Wafaa Mustafa Abduallah and Al Sakib khan Pathan ,”IDPS: An integrated Intrusion

Handling Model for Cloud Computing Environment, International Journal of Computing and Information Technology

( 2012.

1. Ms. Parag K Shelke , Ms. Sneha Sontakke and Dr. A. D. Gawande , “Intrusion Detection System for Cloud Computing”,

International Journal of Scientific & Technology Research,Vol.1,ISSN 2277 8616, 2012.