INTRUSION DETECTION SYSTEM BASED ON FAST LEARNING NETWORK, PARTICLE SWARM OPTIMIZATION AND FUZZY LOGIC

Prajwal Gaikwad, Revati Shelke, TejinderSingh Ramgadiya, Prachi Shinde,

Vivek Shelake ,Shravni Jadhav

Computer Department, All India Shri Shivaji Memorial Society's Institute of Information Technology,Pune, Maharashtra, India

Abstract : System security is a crucial demand now-a-days for huge organizations. The Intrusion Detection frameworks (IDS) are going to be valuable thing for reassurance against attacks that are frequently changing in size and complexity. With information's privacy and accessibility, it must be solid, simple to oversee and with low maintenance cost. Different adjustments are being connected to IDS consistently to recognize new attacks and handle them. This paper proposes a Fuzzy Genetic Algorithm (FGA) for finding the intruders. The FGA framework is a fluffy classifier, whose information base is demonstrated as a fluffy administer, for example, "if-then" and enhanced by a hereditary calculation. The system is tried on the standard KDD'99 dataset. The outcomes are allowing us to show the advantages of the proposed approach.

Keywords— genetic algorithm, fuzzy logic, classification, intrusion detection, KDD'99 data set.

I. INTRODUCTION

An intrusion detection system (IDS) is a system that supervise network traffic for untrusted activities and issues an alert when such activities are discovered. Intrusion Detection System (IDS) has two main types Host Base Intrusion Detection System (HIDS) and Network base Intrusion Detection System (NIDS). Now-a-days system security foundation depends on System Interruption Recognition Framework. It is infeasible to stop interruption attacks, so you should be prepared to handle them. IDS is a cautious component whose main role is to keep work relentlessly considering every possible attack on a framework. Interruption recognition is a procedure used to distinguish suspicious movement both at host and network level. Two principle intrusion detection methods are abnormality identification and anomaly detection. The anomaly identification model illustrates the typical behaviour of a client to recognize this irregular or unusual activity.

Due to the uplift in lifestyle and easy accessibility of internet the amount of data being exchanged has increased. This has made computer system security a global priority. Since it is not realistic to fabricate a plan without any vulnerability, interruption recognition has taken place for a range of analysis. For the most part a gatecrasher is characterized as a framework, project or person who tries to successfully break into a data framework or execute an activity not legitimately permitted. We avoid interruption as any arrangement of procedures that endeavor to trade off the honesty, privacy, or availability of a system asset. The determination of identification procedures that endeavour to trade off the honesty, attentiveness, or accessibility of a PC asset can be suggested as interruption discovery.

An interruption location framework is a gadget or programming application that screens system and/or framework for resentful actions or approach intrusion and produces data to an Administration position. Interruption identification is the procedure of observing the activities happening in a PC framework or organization and breaking them down for trails of likely occurrence, which are awaiting dangers of intrusion of system security arrangements. Fundamentally when an intruder tries to break into a data framework or perform an activity not permitted, we refer to this activity as an interruption. Interruption system may integrate manipulative programming bugs , sniffing unsecured exchange, or misusing the outline defect of express conventions. An Interruption Location Framework is a plan for distinguishing interruptions and reporting them definitely to the best possible power.

II. Literature Review

In the investigation of behind proposed approach, we allude a few speculation it simply portray in deafly Wei Li [1] depict the utilizing hereditary calculation for interruption recognition system distinctive identification procedure. He additionally chipping away at TCP/IP layer for revealing. In the proposed plan is hereditary calculation guideline base framework which including hybrid, change, wellness and collection procedure lastly create the principles for test information. The proposed calculation catches the worldwide semantic data utilizing WordNet. WorldNet is an online lexical database for English dialect. In this methodology he talk about a philosophy of applying hereditary calculation into system interruption discovery procedures. A brief outline of Interruption Identification Framework (IDS), hereditary calculation, and joined discovery procedures are talked about. The framework structural planning is likewise presented. Components influencing the GA are tended to in viewpoint. This achievement of hereditary calculation is extraordinary as it considers just as worldly and spatial data of system relationship amid the encoding of the issue; along these lines, it ought to be more useful for distinguishing proof of system abnormal practices. At long last framework takes a shot at both KDD Glass 99 train and additionally on preparing dataset with the suitable location rate.

Emma[2] Interruption location frameworks use programming figuring system including neural systems and neuro fuzzy systems to sort system conduct and determine what class of assault is being created. Neuro-Fluffy classifiers are utilized for the beginning characterization of the starting system movement. A deduction framework, Fluffy surmising frameworks is further used to choose whether the action is ordinary or angry. Effective IDS frameworks are those skilled of decreasing false positives and deliver high rate assault location. Then again, fluffy surmising framework utilizes human information to create their fluffy standard. Keeping in mind the end goal to present a more exact method for characterizing system exchange, we present the utilization of Hereditary Calculations in blend with ANFIS to improve information arrangement and get the best results. Hereditary calculations utilize an arrangement of hereditary administrators, for example, change, hybrid and choice on current populace to duplicate comparable examples that will be utilized often until a specific paradigm is met. This is a recipient framework essential is utilized to catch every single approaching parcel and store them in an information storage room server. The information is put away as situate of activity streams, with every occurrence being portrayed. This information created are further worked upon to decide their enrollment qualities in view of the diverse parameters characterized in the framework. At that point Preprocessing alludes to the method of separating data about parcel relationship from information and development of new arithmetical components. This procedure orders acquired movement so that enhanced choice can be made on them.

Wenke Lee and Salvatore [3] proposed an Ongoing information mining base IDS, they display an outline of our exploration progressively information mining-based interruption identification frameworks (IDSs). Framework simply centered around issues identified with sending a data mining-based IDS in a continuous domain. Framework portray our ways to deal with location three sorts of issues: precision, productivity, and ease of use. To enhance precision, information mining projects are utilized to break down review information and concentrate includes that can recognize typical exercises from interruptions, they utilize counterfeit oddities alongside ordinary and/or interruption information to deliver more compelling abuse and peculiarity recognition models. To enhance productivity, the computational expenses of elements are investigated and a various model expense based methodology is utilized to create identification models with minimal effort and high exactness.

S. Selvakani [4] speaks to the methodology of interruption identification in system utilizing Hereditary, Fluffy and Apriori Calculation. With broad utilization of web access. The proposed frameworks strategy which is blend of hereditary standard, fluffy guideline and association principle for better result of interruption recognition. Fluffy tenet can arrange system assault information for being a machine learning calculation though hereditary calculation gives best ideal arrangement by finding appropriate fluffy principle and apriori calculation gives best affiliation standard to see assault. Our methodology can be executed on both surely understood dataset i.e. KDD container 1999 and also on own system dataset.IDS will be assessed in states of location rate, recognition speed, false alert rate and assault sorts. GA utilizes normal determination proposal on chromosome like information structure and help in their assessment utilizing choice, recombination and change administrators. In the second stage fluffy Discover Likelihood for every discovery guideline of record to mean genuine positive and genuine negative qualities as appeared in pseudo code underneath. At long last apriori calculation will make the mark base standards for general framework. In this system start with arbitrarily produced populace of chromosomes speaking to all conceivable answer for an issue considered as hopeful arrangement.

Jungwon Kim proposed [5] Towards a manufactured insusceptible framework for system interruption location: an examination of clonal determination with a negative choice operatora new The interruption discovery issue is turning into a testing undertaking because of the expansion of heterogeneous PC systems since the expanded network of PC frameworks gives more prominent access to untouchables and makes it less demanding for interlopers to keep away from ID . Interruption identification frameworks are utilized to identify unapproved access to a PC framework. Various delicate registering based methodologies are being utilized for recognizing system interruptions. This paper introduces a review on interruption location procedures that utilization hereditary calculation approach.

III. System Architecture:-

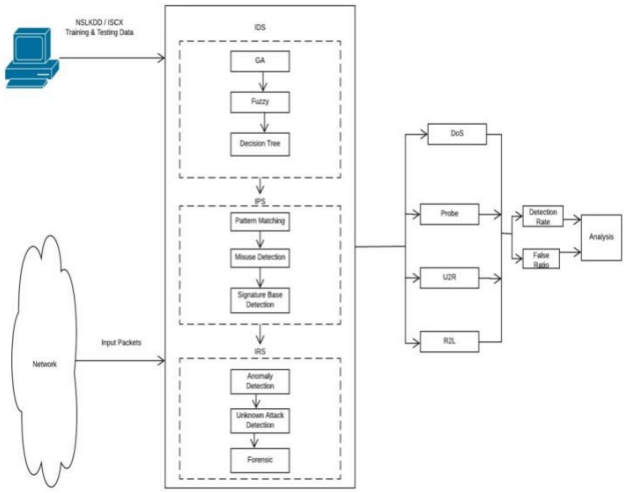

figure 3.1:proposed methodology

In the proposed work we relate the hereditary calculation for GA Rules. In the hereditary development we first create the chromosomes, the collection of these chromosomes we call populace. When populace is delivered apply the single point hybrid and afterward transformation for selecting any random item.

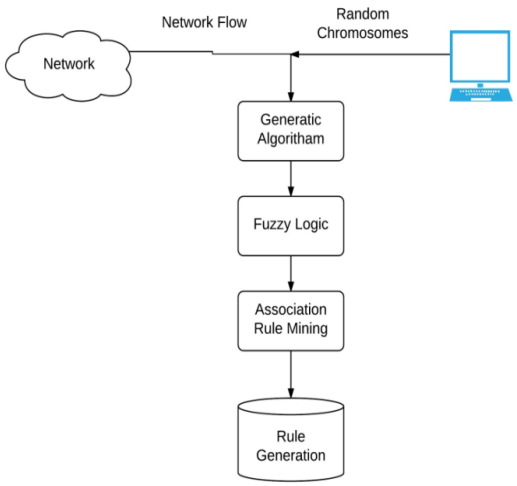

figure 3.2: proposed system architecture (training)

The wellness capacity will mark the wellness estimations of every chromosomes and apply the clustering rule for top principles development. At the point when variety is finished then GA will close and we will get the GA result at last. In second stage we represent the fuzzy validation, based on likelihood capacity. Here the collective will establish base on every characteristic qualities for ordinary interruption base rule.

In the third stage affiliation methodology will create the continuous principles for better identification. Help of these frameworks can effectively recognize the both attacks like inward or outer. The figure 1 demonstrates the general framework structure and information stream.

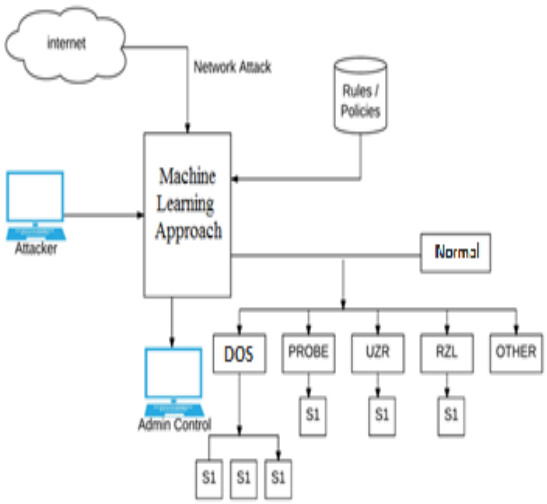

figure 3.3: proposed system architecture (testing)

IV. Algorithms

4.1 Genetic Algorithm

Input: provide the population as p[], number of variations for GA as v, Crossover rate cr, Mutation Rate mr, selection probability sp.

Output: Classification all rules with normal and intrusion pool

Step1: Initialize the random population p.count() for GA.

Step2: Call procedure of crossover

Generate first crossover point and second crossover point

Generate two child chromosomes using crossover points

Step3: Collect all population of crossover pop=sum(p[] +p[]")

Step4: For each (ch in crossover pop)

Generate random r

If(r>mr)

Flip the ch[i] ← Random(i)

Else

Ch[i] ← ch[i]

Step5: Calculate fitness as F= corrected + in corrected / F(x).

End for

Step6: Select all best chromosomes base on sp and remove worse fit chromosome.

Go to step no 2 when (v!=true).

Else terminate GA

4.2. Fuzzy Algorithm [6, 7]

Input : attribute value as from intrusion pool data and Threshold value for similarity

Output: Generate Fuzzy rules for normal as well as anomally.

Assumptions:

1) Consider the 6 parameters for network intrusion are assumed which form the intrusion pool

2) The continuation of trained normal data set (in the experiment accomplish, we have assumed the data of one timing is chosen as the normal trained set)

Step1: Identify and collect relevant data from all intrusion pool.

Step2:Convert the quantitative feature of the data in step 1 into fuzzy sets

Step 3: Define membership function for fuzzy variable

Step 4: Apply probability function to identify the best set of rules.

Step 5: For each of the rules identified in the step 4 do

a) Apply the fuzzy association rule algorithm to mine the correlation among them

b) Apply fuzzy frequency algorithm to mine sequential patterns

Step 6: For each test case generate new patterns using the fuzzy association algorithm for same parameters

Step 7: For each new pattern, compare it with normal patterns created by Training data for similarity

Step 8: IF the similarity > the threshold value

Then report "anomaly" rule pattern

Else normal rule pattern

4.3. Pattern Matching Algorithm for sub attack classification

Input: network connection N which is belongs from KDD set or network adapter, network rules R, Threshold T.

Output: Classify all network instances with label

Step 1: Collect all data traffic from network adapter NF[] as test record.

Step 2: For each (K in R)

For each (items in K=NF[])

Step 3: a[]=split(K);

b[]= split(NF)

Step 4 : Score= CalcSim(a[i],b[i])

Step 5: if (Score > T)

Step 6: match fingerprint to Rules sub type attack

Step 7: define each attack with sub attack SK.

Step 8: Display each NF with SK.

V. Results

We carry out two tests. In the first test, we apply genetic calculation to group typical system information and attack. At that point, we specify identification rate for KDD99 dataset. We characterize them into two classes which are normal and anomaly attacks. In the second test, we utilized the fuzzy calculation to arrange sorts of attacks in the online continuous sniffer dataset.

table 5.1: result analysis

	DOS	Probe	U2R	R2L
TP	98%	96%	70%	94%
FN	2%	4%	30%	6%

figure 5.1: graphical representation of result

VI. Conclusion

In the proposed system we discuss about a method, intrusion detection system using fuzzy genetic algorithm. First stage comprises of genetic algorithm technique for the selection of best chromosome from random population with fitness value. We use fitness function for chromosome fitness calculation. In second stage we apply the fuzzy algorithm on genetic rules it will create a best association rules for testing, finally pattern matching algorithm is used for absolute detection rate with attack types ratio.

Proposed research work also perform the better detection, on the basis of Genetic algorithm implementation we can achieve better detection rate for known attacks as well as unknown attacks. In future work we can minimize the computation time complexity of genetic algorithm.

VII. References

[1] Kagan Tumer a, Adrian K. Agogino(2008), Ensemble Clustering With Voting Active Clusters, Elsevier
[2] Li Zheng Tao Li, Chris Ding(2010),Hierarchical Ensemble Clustering, Ieee International Conference On Data Mining
[3] Sandro Vega-Pons, Jos Ruiz-Shulclopery(2011), A Survey Of Clustering Ensemble Algorithms, International Journal Of Pattern Recognition And Artificial Intelligence Vol. 25
[4] Kaur, P.(2013),Adaptive Intrusion Detection Based On K-Svmeans Algorithm (Doctoral Dissertation, Thapar University Patiala).
[5] Wagh Sk, Kolhe Sr., Effective Semi-Supervised Approach Towards Intrusion Detection System Using Machine Learning Techniques, International Journal Of Electronic Security And Digital Forensics, 7(3):290-304, 2015.
[6] Naila Belhadj Aissa, Mohamed Guerroumi , A Genetic Clustering Technique For Anomaly-Based Intrusion Detection Systems, Ieee, 2015. Networking (Icoin), 2013 International Conference On, Pages 1–5, 2013.
[7] Basant Subba , Santosh Biswas, Sushanta Karmakar ,A Neural Network Based System For Intrusion Detection And Attack Classification, Ieee, 2016.
[8] Wagh Sk, Pachghare Vk, Kolhe Sr, Survey On Intrusion Detection System Using Machine Learning Techniques, International Journal Of Computer Applications. 1;78(16). Jan 2013.
[9] Mohammed A. Ambusaidi Et. Al., Building An Intrusion Detection System Using a Filter-Based Feature Selection Algorithm , Ieee Transactions On Computers, Vol., No , November 2014.
[10] Fatemeh Barani , A Hybrid Approach For Dynamic Intrusion Detection In Ad Hoc Networks Using Genetic Algorithm And Artificial Immune System, Ieee , 2014.
[11] Levent Koc And Alan D. Carswell , Network Intrusion Detection Using a Hnb Binary Classifier In Ieee 2015
[12] Eman Abd Ei Raoof Abas ,Artificial Immune System Based Intrusion Detection: Anomaly Detection And Feature Selection Ieee 2015.