Intrusion Detection System using Memtic Algorithm Supporting with Genetic and Decision Tree Algorithms

[1]K.P.Kaliyamurthie , [2]D,Parameswari , [3]DR. R.M. Suresh

[1] Assistant Professor, Dept of IT, Bharath University.
Chennai,Tamil Nadu-600073.

[2] Assistant Professor, Dept of MCA,Jerusalem College of Engineering.
Chennai,Tamil Nadu-600100.

[3]. Professor & Head, Dept of CSE,RMD Engineering College.
Chennai, Tamil Nadu – 601206.

Abstract

This paper has proposed a technique of combining Decision Tree, Genetic Algorithm, DT-GA and Memtic algorithm to find more accurate models for fitting the behavior of network intrusion detection system. We simulate this sort of integrated algorithm and the results obtained with encouragement to further work.

Keywords. ,DT,GA,memtic,hostbased IDS, network based IDS.

Introduction

Computer networks are usually protected by anti-virus software, firewall, and encryption, secure network protocols, password protection etc. Since it has been proven that a potential attacker can always find a way to attack a network. So we need additional support that would detect this type of security breaches. These systems are known as IDS and are placed inside the protected network, looking for potential threats in network traffic and or audit data recorded by host.

There are two different types of intrusion detection (i) misuse detection ii) anomaly intrusion detection.[6]

Misuse intrusion detection uses well-defined patterns of the attack that exploit weakness in system and application software to identify the intrusions. Anomaly intrusion detection system identifies abnormal behavior of the network at a certain period.

There are basically two main types of IDS being used today: Network based (a packet monitor), and Host based (looking for instance at system logs for evidence of malicious or suspicious application activity in real time).

A network based IDS usually consists of a network appliance (or sensor) with a Network Interface Card (NIC) operating in promiscuous mode and a separate management interface. The IDS is placed along a network segment or boundary and monitors all traffic on that segment.

A host based IDS requires small programs to be installed on individual systems to be monitored. The agents monitor the operating system and write data to log files and/or trigger alarms. Host-based IDS can only monitor the individual host systems on which the agents are installed; it doesn't monitor the entire network.

Data mining approaches can be used to extract features and form rules. Data mining is used for analyzing data in order to find rules and patterns describing the characteristics properties of the data. Intrusion detection system process to monitor large amount of data.

Many data mining techniques are available like association rule mining. Decision tree was the best choice because the data set was huge and multi dimensional. In decision tree model, parent class is formed by more than one sub classes. The decision trees classify the model into tow classes as 'normal' and 'attack'. Decision tree methods can be useful for classifying log data and detecting intrusions.[4]

Genetic Algorithm has been used in different ways in IDS.GA to generate artificial intelligence rules for IDS. Our network connection and its related behavior can be translated to represent a rule to judge whether or not a real-time connection is considered an intrusion. These rules can be modeled as chromosomes inside the population. The population evolves until the evaluation criteria are met. The generated rule set can be used as knowledge inside the ID for judging whether the network connection and related behaviors are potential intrusion.

In this paper, we proposed DT based on J48, GA and the combination of DT and GA for network intrusion detection. Since GA is one of the most powerful search technique to find the approximate solution to combinatorial optimization problems. We used KDD Cup'99 data set generated by MIT Lincoln Laboratory has been used in our experiment [5].

Decision Tree

Decision Tree induction is one of the classification algorithms in Data mining. The classification algorithm is inductively learned to construct a model from the reclassified data set (Briement et al.,1984). In our approach we have used the J48 algorithm for decision tree to audit data.

J48 Algorithm

J48 implements Quinlans c4.5 algorithm for generating a pruned or unpruned C4.5 decision tree. The decision tree generated by J48 can be used for classification. J48 builds decision tree from a set of leveled training data using the concept of information entropy. It uses the fact that each attribute of the data can be used to make a decision by splitting the data into smaller subsets. J48 examines the normalized information gain that results from choosing an attribute for splitting the data. To make the decision, the attribute with the highest normalized information gain is used. Then the algorithm recurs on the smaller subsets. The splitting procedure stops if all instances in a subset belong to the same class. Then a leaf node is created in the decision tree telling to choose that class. In this case J48 creates a decision node higher up in the tree using the expected value of the class .It handle both continuous and discrete attributes, training data with missing attribute value and attribute with differing cost. Further, it provides an option for pruning trees after creation.

Genetic Algorithm

Genetic Algorithm were formerly introduced in the United states in the 1970's by John Holland at University of Michigan. The continuing performance improvement of computational systems has made them attractive for some types of optimization. Genetic Algorithm is very powerful search technique used in computer science based evolution principle. Genetic Algorithm work very well on continuous and discrete combinatorial problems.

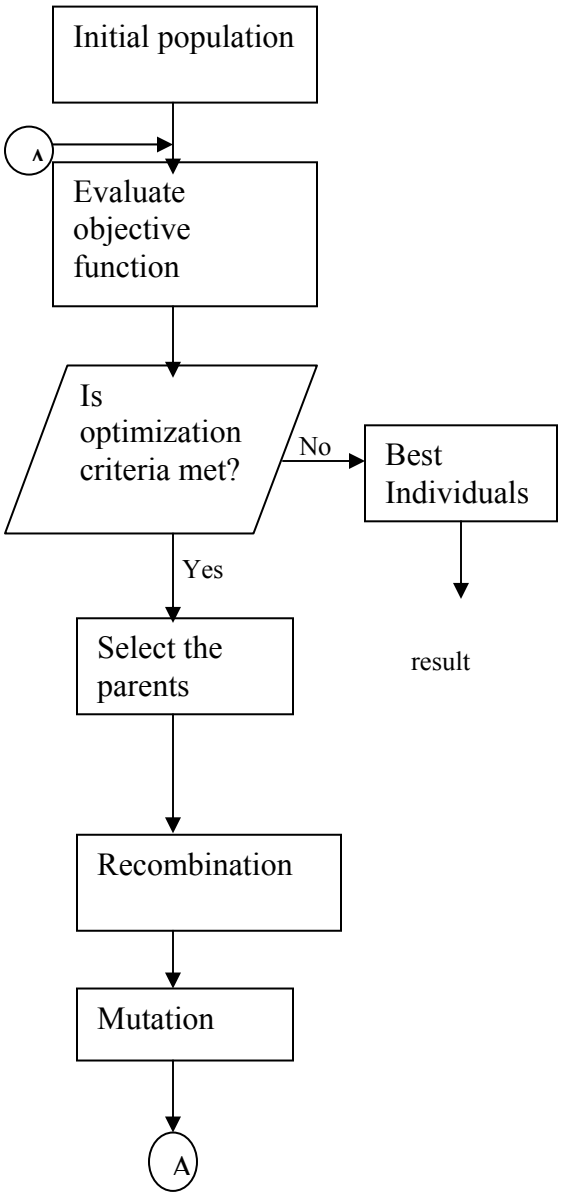
Process of Genetic Algorithm

Genetic Algorithm uses an evolution and natural selection that uses a chromosome-like data structure and evolve the chromosomes using selection, recombination, and mutation operators.

The process usually begins with randomly generated population of chromosomes, which represent all possible solution of a problem that are considered candidate solutions. Different positions of each chromosome are encoded as bits, characters or numbers. These positions could be referred to as genes.

An evolution function is used to calculate the goodness of each chromosome according to the desired solution; this function is

known as "fitness function". During evaluation, two basic operators, cross over and mutation are used to simulate the natural reproduction and mutation of species. The selection of chromosomes for survival and combination is biased towards the fittest chromosomes.



Fig(1) process of GA

This figure shows the structure of Genetic Algorithm. Starting by a random generation of initial population. There are four types of encoding chromosomes Binary Encoding represents chromosome gene in 0's and 1's. Permutation encoding represents chromosomes gene by integer values. Real valued Encoding represents chromosome gene by the real values of the events in the problem. Tree encoding represent chromosome in a tree form for some object.

Create a random initial state:

An initial state is created from a random selection of solutions.

Evaluate Fitness

A value of fitness is assigned to each solution (chromosomes) depending on how close it actually is to solving the problem. The fitness value evaluates the performance of each individual in the population.

Crossover

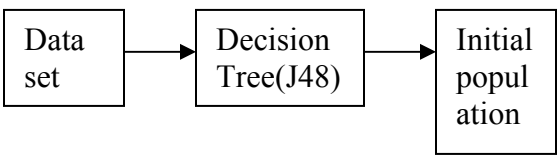
Those chromosomes with a higher fitness value are likely to reproduce offspring. The offspring is a product of the combination of genes.

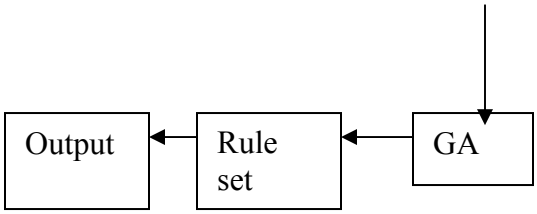
Next generation

If the new generation contains a solution that produces an output that is close enough or equal to the desired answer then the problem has been solved. If this is not the case, then the generation will go through the same process as their parents did. This will continue until a solution is reached.

Experiment setup and performance evolution

We used KDD Cup 1999 intrusion detection data set .This set defined the features between normal connections from attacks We used four types of attacks out of 24 attacks. There are Deniel of service (DOS), Remote to used(R2L),User to root(U2R) and probing.





Fig(2) Architecture of DT and GA applying intrusion detection.

Decision Tree

The decision tree is partitioned into two classes of normal and Attack. Our object is to separate normal and attack pattern according to the attack class. This process is repeated for all the classes. The classifier was constructed using the training data and testing data.

Applying Genetic Algorithm

The GA approach work with two two modules with different stage. In the training state, GA used in an offline environment. In the ID stage the generated rules are used to classify incoming network connection in the real-time environment. The attributes are protocol, source and destination port IP and attack-name.

It generates initial population, sets the default parameter and audits the data. The initial population is evolved by the number of generation. Quality of the rules is initially calculated, then a number of best-fit rules are selected and finally GA operators are applied to all generations. Offline training system and online detection system that uses the generated rules to classify incoming network connections in real-time environment.

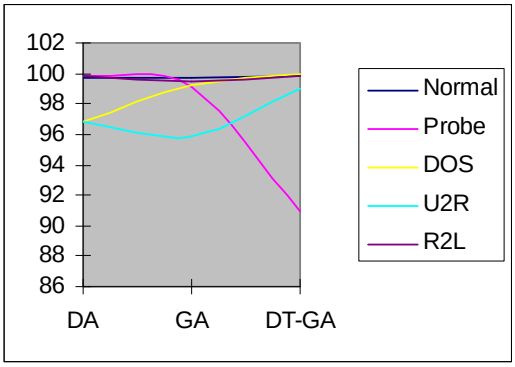
DT and GA

The data sets were first passed through the DT and the node information was generated. Training and testing data along with the node information is given to the GA.

Performance comparison of DT,GA and DT-GA.

Attack	Accuracy

Type	Decision Tree	GA	DT-GA
Normal	99.66	99.66	99.80
Probe	99.87	99.12	99.90
DOS	96.85	99.20	99.92
U2R	96.85	95.80	98.94
R2L	99.86	99.42	99.86



3.MEMETIC ALGORTHIM

Combining global and local search is a strategy used by many successful hy brid optimization approaches.

Memetic Algorithms (MAs) are Evolutionary Algorithms (EAs) that apply some sort of local search to further imp rove the fitness of individuals in the population. Memetic Algorithms have been shown to be very effective in solving many hard combinatorial optimization problems. The app roach combines a hierarchical design technique, Genetic Algorithms, constructive techniques and advanced local search to solve VLSI circuit it Layout in the form of circuit routing. Results obtained indicate that M emetic Algorithms based on local search, clustering and good initial solutions imp rove solution for the VLSI circuit routing problem.

The traditional app roach in routing is to construct an initial Solution by using constructive heuristic algorithms. A final solution is then produced by using iterative improvement techniques where a modification is usually accepted if a reduction in cost occurs, otherwise it is rejected. Constructive heuristic algorithms Produce an initial solution from scratch. It takes a negligible amount of computation time Compared to iterative improvement algorithms and provides a good starting point for them (SM 91). However, the solution

generated by constructive algorithms may be far from optimal. Thus, an iterative improvement algorithm is performed next to imp rove the solution.

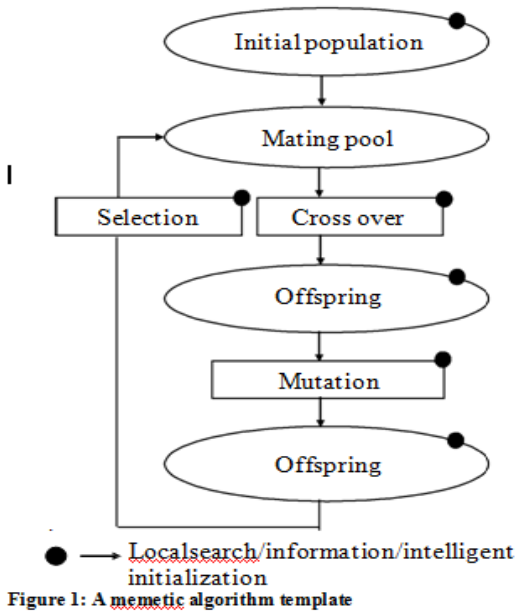


Figure 1: A memetic algorithm template

Local search/information/intelligent initialization

Figure 1: A memetic algorithm template

Although iterative improvement algorithms can produce a good final solution, the computation time of such algorithms is also large. Therefore, a hierarchical approach in the form of multilevel clustering is utilized to reduce the complexity of the search sp ace. A bottom-up technique gradually clusters cells at several levels o f th e hierarchy. At the top level a Genetic Algorithm is app lied where several good initial solutions are injected to the population

A local search technique with dynamic hill climbing cap ability is app lied to the chromosomes to enhance their quality. The system tackles some of the hard constraints imposed on the problem with intermediate relaxation mechanism to further enhance the solution quality. The pseudo-code for memetic algorithm is presented given below.

```
Begin
For j: =1 to μ do
    I: =gen
 generate
 solution
 (); I:
 =local
 search (I);
 Add individual I to P;
endfor
Repeat
 For i: = 1 to p cross. μ do
    Select two parents I_A,I_b Є P randomly;
    I_c:
    =recom
    bine
    (I_a, Ib);
    I_c: =
    local-
    search(
    I_c);
    Add individual I_c to P";
     endfor;
              I_m : =local-search (I_m )
              ; Add individual I_m to P";

     P: =select (P υ P")
If conver ged (P) then P: =local-
search (mutate (P)); Until
terminate=true;
end
```

Conclusion

In this paper, we implemented DT and GA approach for detecting known and novel attacks on the network, This approach proven their efficiency in both generalization and detection of new attacks. The accuracy of DT-GA gives better performance when compared to the accuracy of DT and GA.

We have successfully included the Memtic algorithms with the above DT-GA algorithm as a combination and hence paved the way for looking at the network Intrusion Detection System with a new perception.

Acknowledgement : We thank our management for the support and encouragement for carrying out this work.

Reference

1] Ben Coppin, Artificial Intelligence Illuminated,2005.

2] Graham, Robert,"Network Intrusion detection systems", 2002.

3] Tamas Abraham "Intrusion detection using Data mining Techniques".

4] I.H.Witten and Efrank"Data mining: Practical machine learning tools and techniques with java implementations", Academic Press, Morgan Kauffman.

5] KDD Cup'99 Intrusion Detection data sets. Data available at:http//kdd.ics.uci.edu/database/kddcup 1999/kdd cup99.html,1999.

6] Sinclair,chris,Lynpierce and Sara Matznel,1999."An application of Machine learning to network intrusion detection", In proceedings of 1999 Annual Computer Society Applications Conference.pp-371-377.

7] S.Kumar and E.H. Spaffort," a Software Architecture to support misuse Intrusion Detection", Technical Report, 1995.

8] W.Li. "Using Genetic Algorithm for networkIntrusionDetection",Proceedings of the United States Department of Energy Cyber Security Group",2004.

9] A.Abraham and C.Grosan "Evolving Intrusion Detection Systems",Studies in computational intelligence(SCI),2006.

10] Banzhaf N.Nordin P,Killer E.R,Francone FD, "Genetic Programming: An introduction on the automatic evolution of computer programs and its applications",1998.

11] K.Scarfone and P.Mell,"Guide to intrusion detection and prevention systems", National institute of standards and Technology NIST, Computer Society, 2007.

12]. Natalio Krasnogor and Jim Smith, "A Tutorial for comp etent memetic algorithms: model, taxonomy and design issues", IEEE transaction on evolutionary computation, vol a, no. b, CCC 200d, 2005.

13]. Hawkiareibi, School of Engineering, University of Guelph, Guelph, Ontario, N1G2W1,Canada, Zhen Yan g , School of Engineer in g, U niversity of Guelph, Guelph, Ontario, N1G2W1,Canada "Effective M emetic Algorithms for VLSI Design

Mr.K.P.Kaliyamurthie received B.E (CSE) degree from Bharathidasan University in 1990 and M.Tech(CSE) from Bharath University in 2007. I am pursuing Ph.D Degree in Computer Science and Engineering since 2008 from Bharath University,Chennai. I am having 20 years of teaching experience ,currently I am working as a Asst.Prof in Bharath University,Chennai. I have published several papers on national , international journal and conferences.

Ms.D.Parameswari received B.Sc (CSE) degree from Bharathidasan University in 1994 and M.Sc(CS) from Bharathidasan University in 1996,MPhil (CS)2003 and MCA degree 2009 from Manonmaniam Sundarnar university.I am pursuing Ph.D Degree in Computer in Mother Teresa Womens University, Kodaikanal.I am having 15 years of teaching experience ,currently I am working as a Asst.Prof in Jerusalem College of Engineering,Chennai. I have published several papers on national , international journal and conferences.

Dr. R.M.Suresh , a versatile and active personality is known for his enterprising spirit. After completing his B.E in CSE from Bharathidasan University in 1989 and M.Tech in Computer Engineering from Mysore University in 1992 he went ahead to complete his PhD in Computer science and Engineering in the year 2001. He is in teaching profession for the past 21 years. He initially served as a lecturer in CSE Department of Manonmaniam Sundaranar University for a decade and moved ahead to RMK group of institution. He is currently serving also as Associate Dean (R & D) and Vice Principal of this College.

He has published 10 papers in International journals and presented 85 papers in refereed national & International Conferences and has written a book in Tamil "C – Programming" supported by Tamil Nadu Govt. He has conducted 5 summer/winter schools and more

than 20 conferences and workshops. He is currently guiding 12 PhD scholars and has already guided more than 24 M.Phil scholars, 26 M.Es, in the preparation of Thesis. He is a member of IEEE, CSI, ISTE, INFITT; IASTE etc .His contribution to the academia is quite significant