AN INTRUSION DETECTION MODEL BASED ON FUZZY MEMBERSHIP FUNCTION USING GNP

Mahadevappa Immannavar[1], Prasad Pujar[2], Manjunath Suryavanshi[3]

[1]*Dept. Of Computer Science and Engineering, K. L. E College of Engineering & Technology, Chikodi-591201*
[2]*Dept. Of Computer Science and Engineering, K.L.S Gogte Institute Of Technology, Belagavi-590008*
[3]*Dept. Of Computer Science and Engineering, K. L. E College of Engineering & Technology, Chikodi-591201, Karnataka, India*

Abstract

As the Internet facilities increasing over the world, threats, attacks or intrusions over the Internet are also increasing. Therefore, an intrusion detection model is required to detect intrusion that going to threaten CIA of internet resources. A GNP based fuzzy membership function is much more suitable for identifying such kind of intrusions. A GNP which is a combination of GA and GP applied to extract association rules. A combined GNP-fuzzy membership method would help us to extract important association rules from DARPA 98/99 dataset rather than all rules from DARPA 98/99 dataset. Then the extracted association rules would be updated using genetic operations and also stored into rule pool. In classification, association rules will be classified as normal or intrusion based on calculated match degree. The classified association rules will be stored separately in two different rule pools. Normal rules in normal rule pool and intrusion rules in intrusion rule pool. For the new data match degree will be calculated based on available normal rules and intrusion rules. Then this calculated match degree will help us to identify whether the new data normal or intrusion.

Keywords: *Fuzzy membership function, Genetic network programming, Genetic algorithm, DARPA 98/99 dataset and Intrusion detection.*

---***---

1. INTRODUCTION

For the past decade rapid growth in computer networks security has become an important issue for computer users. Day by day services from the Internet applications over Internet such as net- banking, online shopping, trading stocks etc are increasing. In order to get proper services from these applications, security mainly required. Meanwhile, this process of getting services form applications may damages or crises by intrusions or attacks. Therefore, intrusion detection plays a main role in network security to order identify threats or attacks and to take proper action for the identified attack or threat.

1.1 Intrusion Detection

It is of process of identifying threats or attacks that going to threaten integrity, confidentiality, or the availability of network resources. Intrusion detection can be done in either of two ways namely manual detection or automatic detection. Manual threat detection would detect threats by examining stored files known as log files. In case of automatic intrusion detection, intrusion detection software would be installed on the required machine for doing the same.

Anomaly detection and misuse detection are the main two types of intrusion/attack detection techniques. Misuse detection uses previous attacks to identify intrusions, whereas anomaly detection uses normal behaviors to detect unknown attacks. Below figure show the architecture of misuse detection system. Misuse detection based on stored rules identifies intrusions. That is it compares generated new rule of entered connection with the existing rules, if matching is found then it respond to the administrator as intrusion through alarm. It keeps on updates system profile which is going contain rules that identifies attacks. That is, it updates existing rules as well as adds new rules which are going to generate from newly entered data. The process of updating existing rules is known as modification.

Misuse detection would detect all kind of attacks. That itself has drawback of misuse detection system. That is, it would detect only the standard attacks that going break up security attacks.

1.2 Association Rule Mining

DM normally refers to the process of fetching needed rules from large data container to perform other actions. The rapid development in data mining allowed building much more methods suitable for intrusion-identification problems. ID can be considered as a classification approach that is it will classify audit record as either normal or intrusion. An ARM is one of the most commonly used methods to extract association rules which exist among the set of attributes within the dataset. Association rule mining itself shows relationship between the set of attributes. An association rule A→B, where A and B are the set of attributes means that if someone record satisfies A, it also satisfies B.

The most commonly used rule mining algorithm is apriori algorithm. Actually apriori algorithm is used or exists for extracting frequent itemsets. Frequent itemsets intern generates association rules. Important measures of association rules are support count and confidence. This algorithm would suffer from a lot of problems. One among them is it could not handle large dataset.

2. RELATED WORK

2.1 GNP'S Basic Structure

GNP itself one of the evolutionary technique, which using graph structures rather than strings and trees. GNP contains-SN, JN and PN.

At the beginning every record passes through start node. Judgment nodes say J_1, J_2, $J_{3.....}$ J_m makes decision in order to find next node in the GNP structure. Processing nodes say P_1, P_2, P_3 ….. P_n would process n functions.

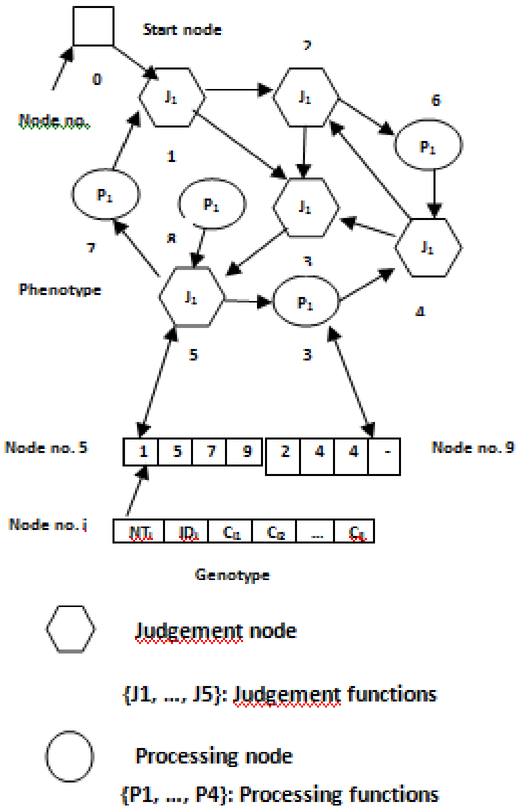

Fig 1 GNP's Basic structure for individual

Actual actions for the processing nodes would define in advance and stored in the storage. Once the GNP is start up, every record passes nodes to generate association rules. In the GNP individual, NT_i would represents node type as per below code

0- For SN
1- For JN and
2- For PN.

Within the GNP individual, ID_i would represents the id number of node, e.g NT=2 and ID=2 refers P_2.

2.2 Class-Association-Rule Extraction

Let P be a group of attributes/literals/ items and L be the group of records, where each record R is a collection of attributes | R P. A record R contains attributes X in P, if X P.

An implication X=>Y is called an association rule. Let sup(X) =p the number of records holding X in L, sup(Y) =q, and sup(X∪Y) =r. Confidence of X=>Y is defined by Sup(X ∪Y)/Sup(X) =r/p.

Let sup(X) =p, sup(Y)=q, sup(X∪Y)=r, and n represents the records number in the dataset. Then the equation for chi-square test

$$\text{Chi-Square}=\frac{(r-pq)(r-pq)n}{(1-p-q+pq)\,pq} \;\ldots\ldots\; (1)$$

Class association rules satisfies the below defined conditions would be considered as important rules.
$Sup_{min} \leq Sup$
$Conf_{min} \leq Conf$
$\text{Chi-Square}_{min} \leq \text{Chi-Square}$

Minimum values for support, confidence and chi-square are always predefined.

3. CARM BASED ON GNP

3.1 Representation of Association Rules

A judgment node checks attribute value to choose the correct path. Association rules would be represented through the connections occurs between judgment nodes. In case of node transition to generate association rules, for every tuple which satisfies the condition yes arm will be chosen and the next decision will be checked. No arm is selected to proceed with the P_2 to begin the examining of another rule.

3.2 Working Procedure for Generating Association Rules is as Follows

1^{st} record will be fetched from the DB and transformation starts from start node. Then, if yes-arm is selected, then it moves to next judgment node. Otherwise it moves to the next processing node to generate another rule. This process is repeatedly performed until it reaches last node P_n.

3.3 Calculation of Measurements for Association Rules

Let N be the number of records , p, q, and r be the number of records moving towards yes-side of first judgment node, second judgment node and third judgment node respectively. For example, for the rule (B_1=1) => (Class=1), the support count is p(1)/N & the confidence is p(1)/p. For (B_1=1) and (B_2=1) and (B_3=1) => (Class=1), support is r(1)/N and confidence is r(1)/r. Chi-square value will be calculated by using above equation mention for chi-square.
Based on the chosen minimum values for support, confidence, and chi-square important rules will be extracted

4. FUZZY MEMBERSHIP FUNCTION

4.1 Subattribute Utilization

Subattribute utilization is used to prevent the data loss. It mainly used divide symbolic attribute into several attributes, binary into two attributes with two values such as 0 and 1, and continuous attributes into three attributes such as low, mid, and high. Fuzzy membership function is required to generate fuzzy membership values and values for fuzzy parameters alpha, beta and gamma for the partitioned continuous attributes.

In this case each continuous attribute's value is divided into three linguistic terms say low, mid and high. That is each continuous attribute is divided into three sub attributes with three linguistic values low, mid and high. Fuzzy membership function to each continuous attribute will be predefined. Fuzzy parameters would be calculated as per below representation.

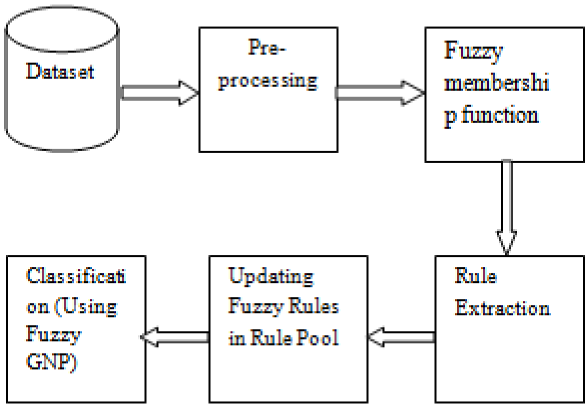

Fig 2 System Architecture

Beta - average value of continuous attribute,
Gamma-highest value of continuous attribute, and
Alpha + Gamma= 2.Beta

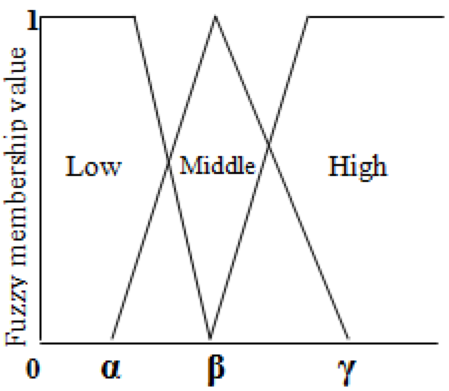

Fig 3 Definition of the fuzzy membership function

Table 1- Sample Database

TID	B1	B2
1	20	800
2	10	600
3	40	400
4	50	200
5	30	1000

For the attribute B1 the fuzzy parameter values alpha=20, beta=30 and gamma=50. For the attribute B2 the fuzzy parameter values alpha=200, beta=600, and gamma=1000.

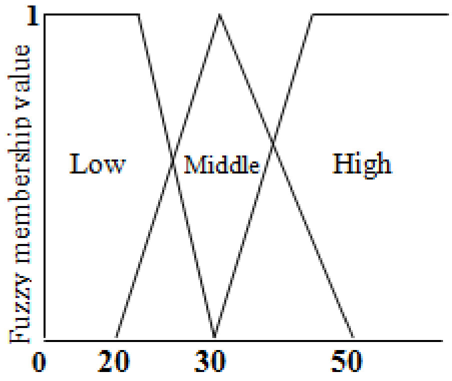

Fig 4 MF for Attribute *B*1

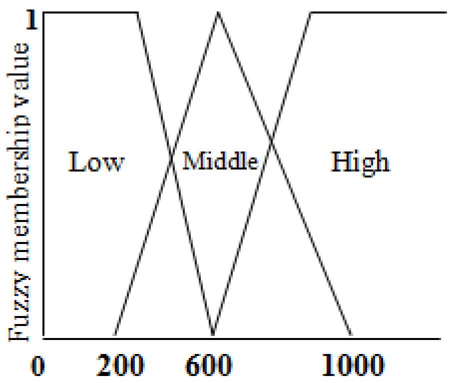

Fig 5 MF for Attribute *B*2

Table 2-DB with FM Values

TID	Attribute B_1			Attribute B_2		
	Low B_{11}	Mid B_{12}	High B_{13}	Low B_{21}	Mid B_{22}	High B_{23}
1	1.0	0	0	0	0.5	0.5
2	1.0	0	0	0	1.0	0
3	0	0.5	0.5	0.5	0.5	0
4	0	0	1.0	1.0	0	0
5	0	1.0	0	0	0	1.0

4.2 Rule Extraction

In this step from the dataset DARPA-98 or DARPA-99 association rules will be extracted for the tuples transferred over the node transition diagram. GNP examines all the tuples to generate required rules. The training dataset help us generate two categories of rules. The generated rules will be stored in their corresponding pools. Separate rule pools will be maintained for each kind of rule. That is one rule pool to hold normal rules and its also known as normal rule pool. Another rule pool will hold intrusion and it's known as intrusion rule pool.

Figure 6 would shows flow diagram of our proposed ID system.

Flow Diagram

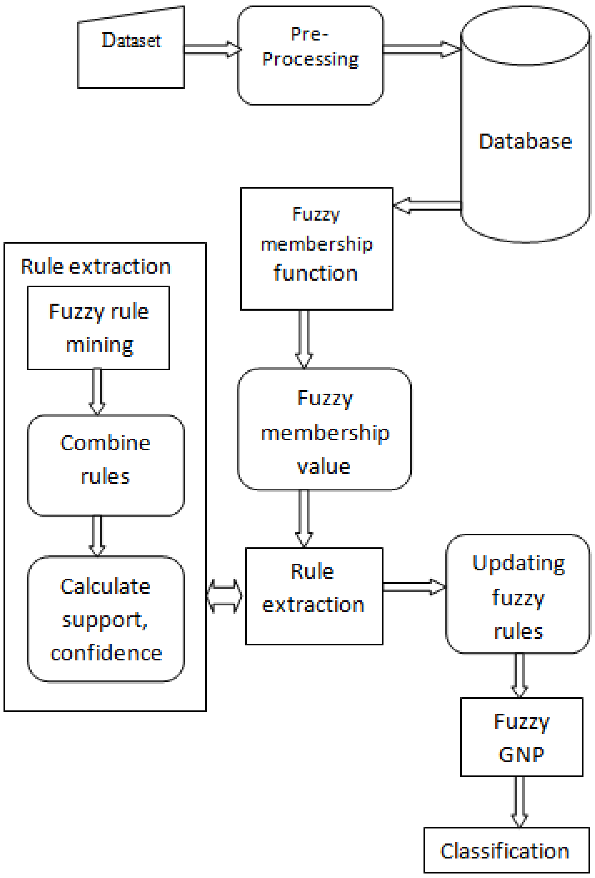

Fig 6 Flow Diagram

4.3 Updating Fuzzy Rules

Fuzzy CARs will be generated and stored into the storage called pool through generations along with their calculated conf, sup and chi-square. If any rule occurs with more sup, conf & chi-square, then it will the same rule in the storage.

4.4 Fitness and Genetic Operation

The fitness of extracted class association rule *r* is defined by:

$$fitness_r = \frac{Numt_c}{Numt} - \frac{Numn_i}{Numn} \text{.........(2)}$$

The fitness value range is [–1, 1].

In every generation, individuals would replace through genetic operations in order to generate more class-association rules. Genetic operations are always- selection, mutation and cross over and these many we specify through genetic algorithm.

4.5 Classification

Here we calculate MATCH*n* (*d*new), and MATCH*i* (*d*new). If MATCH*i* (*d*new) ≤ MATCH*n* (*d*new), new connection data *d*new will be labeled as normal. If MATCH*i* (*d*new)≥MATCH*n* (*d*new), new data *d*new will be labeled as intrusion.

5. RESULTS

In our proposed method both normal and intrusion rules will be extracted from DARPA-98 and DARPA-99. Below table shows the number of normal rules extracted from the DARPA 98 dataset for the given minimum support count, minimum confidence and minimum chi-square.

Table 3- Number of Normal Rules Extracted from DARPA-98

Sl. No	Min Sup	Min Confidence	Min Chi-Square	No of extracted normal rules
1	0.1	0.1	0.1	66
2	0.2	0.2	0.2	60
3	0.3	0.3	0.3	54
4	0.4	0.4	0.4	54
5	0.5	0.5	0.5	54
6	0.6	0.6	0.6	54
7	0.7	0.7	0.7	54
8	0.8	0.8	0.8	54
9	0.9	0.9	0.9	54
10	1.0	1.0	1.0	54

Below table shows the number of intrusion rules extracted from the DARPA 98 dataset for the given minimum support count, minimum confidence and minimum chi-square.

Table 4- Number of Intrusion Rules Extracted from DARPA-98

Sl. No	Min Sup	Min Confidence	Min Chi-Square	No of extracted intrusion rules
1	0.1	0.1	0.1	44
2	0.2	0.2	0.2	32
3	0.3	0.3	0.3	32
4	0.4	0.4	0.4	32
5	0.5	0.5	0.5	32

6	0.6	0.6	0.6	32
7	0.7	0.7	0.7	32
8	0.8	0.8	0.8	31
9	0.9	0.9	0.9	31
10	1.0	1.0	1.0	31

6. CONCLUSION & FUTURE ENHANCEMENT

6.1 Conclusion

Now we conclude our proposed system with the following details-

- It handles both discrete and continuous attributes.
- It can be flexibly applied to any kind of attacks.
- Fitness function help us to retrieve much rules from DARPA-98/99
- GNP would perform effective rule mining.

6.2 Future Enhancement

- In future, we will focus on various specific distribution methods such as poison distribution, binomial distribution and so on.

REFERENCES

Papers

[1] B. Uppalaiah, K. Anand B. Narasimha, S. Swaraj, T. Bharat, *"Genetic Algorithm Approach to Intrusion Detection System"*, IJCST Vol. 3, pp. 156-160, Jan-2012

[2] Shingo Mabu, Ci Chen, Nannan Lu, Kaoru Shimada, and Kotaro Hirasawa, "*An Intrusion-Detection Model Based on Fuzzy Class-Association-Rule Mining Using Genetic Network Programming"* IEEE Transactions On Systems, Man, And Cybernetics Vol. 41, No. 1, January 2011

[3] M. Crosbie and G. Spafford, "*Applying genetic programming to intrusion detection"* presented at the AAAI Fall Symp. Series, AAAI Press, Menlo Park, CA, Tech. Rep. FS-95-01, 1995

[4] Mohammad Sazzadul Hoque, Md. Abdul Mukit and Md. Abu Naser, "*An Implementation of Intrusion Detection System Using Genetic Algorithm*", International Journal of Network Security & Its Applications (IJNSA), Vol.4, No.2, March 2012

[5] Swati Dhopte and N. Z. Tarapore, " *Design of Intrusion Detection System using Fuzzy Class-Association Rule Mining based on Genetic Algorithm*", International Journal of Computer Applications (0975 – 8887) Volume 53– No.14, September 2012.

[6] K. Shimada, K. Hirasawa, and J. Hu, "*Genetic network programming with acquisition mechanisms of association rules*," *J. Adv. Comput. Intell. Intell. Inf.*, vol. 10, no. 1, pp. 102–111, 2006.

[7] Luo J., "*Integrating fuzzy logic with data mining methods for intrusion detection,*" Master's Thesis, Department of Computer Science, Mississippi State University, Starkville, MS, 1999.

[8] Semaray J., Edmonds J., and Papa M., *"Applying data mining of fuzzy association rules to network intrusion detection*," presented at the IEEE Workshop Information, United States Military Academy, West Point, NY, 2006.

[9] Agrawal R. and Srikant R., "*Fast algorithms for mining association rules,*" in Proceeding 20th VLDB Conference, Santiago, Chile, pp. 487–499, 1994.

[10] Shetty M. and Shekokar N., "*Data Mining Techniques for Real Time Intrusion Detection Systems,*" International Journal of Scientific & Engineering Research Volume 3, Issue 4, April 2012.

[11] Mabu S., Chen C., Shimada K., "*An Intrusion-Detection Model Based on Fuzzy Class-Association-Rule Mining Using Genetic Network Programming*," IEEE Transactions Systems, Man, Cybernetics C, Application and Reviews, volume 41, number 1, pp. 130–139, January 2011.

Books

[12] Pang-Ning Tan, Michael Steinbach and Vipin Kumar, "*Introduction to Data Mining*", Addison-Wesley publication, pp-327-386, 2006.

[13] Jiawei Han and Micheline Kamber, "*Data Mining: Concepts and Techniques",* Morgan Kaufmann Publishers, pp-243-248, March 2006**.**

[14] Ramez Elmasri and Shamkant B. Navathe, *"Fundmentals Of Database Systems",* Pearson publication, fifth edition, pp. 963-994, 2011.

Websites

[15] Kddcup 1999data [Online]. Available: kdd.ics.uci.edu/databases/kddcup99/kddcup99 .html.

[16] Darpa Intrusion Detection datasets [Online]. Available: www.ll.mit.edu/mission/ communication/communications/ist/corpra/ideval/data/index.html

[17] [http://www.wikipedia.org

BIOGRAPHIES

Mr. Mahadevappa Immannavar received the Bachelor of Engineering degree in Computer Science and Engineering from Basaveshwar Engineering College, Bagalkot, affiliated to VTU, Belagavi, during 2009. He is persuing M.Tech at Gogte Institute of Technology, Belagavi, under Visvesvaraya Technological University, Belagavi. Currently he is working as an assistant professor in Computer Science and Engineering Department of K. L. E. College of Engineering and Technology, Chikodi since from 2010.

Mr. Prasad M. Pujar, received his M.Tech degree in computer science and engineering and currently working as an assistant professor in Computer Science and Engineering Department of K. L. S. Gogte Institute of Technology, Belagavi, affiliated to Visvesvaraya Technological University, Belagavi. He had eight years of teaching experience and two years of research experience.

Mr. Manjunath Suryavanshi received a Bachelor of Engineering degree in Computer Science and Engineering from Gogte Institute of Technology, affiliated to VTU, Belagavi, during the year 2009. He completed his M.Tech. degree in Software Engineering from M. S. Ramaiah Institute of Technology, an autonomous institute affiliated to VTU, Belagavi, during the year 2013. He is working as an assistant professor in Computer Science and Engineering Department of K. L. E. College of Engineering and Technology, Chikodi, since from August-2011.