CHALLENGES IN GENETIC ALGORITHM BASED INTRUSION DETECTION

Dharmendra G. Bhatti, Dr. P. V. Virparia, Dr. Bankim Patel

ABSTRACT

Intrusion detection is the technique of detecting malicious traffic on a network or a device. It is one of the critical network security components against emerging intrusions techniques and attacks. In this paper we present a survey of different intrusion detection approaches. Intrusion Detection Systems based on Genetic Algorithm are currently attracting researchers due to its inherent potential. Intrusion detection faces various challenges like reliably detect malicious activity and perform efficiently to cope with the large amount of network traffic. Here we have analyzed the present research challenges and issues in Genetic Algorithm based intrusion detection. Finally we carry out our experiments based on our sample Genetic Algorithm using KDD Cup 99 data set. The main contribution of the implementation is the understanding of challenges in Genetic Algorithm based intrusion detection.

Keywords: Security, Challenges, Genetic Algorithm, Intrusion Detection.

1. INTRODUCTION

Internet becomes a need for business as well as home users. Along with its great services Internet has also increased attack rate drastically. But in today's communication era we cannot deny Internet services. Even if best security precautions are present successful attacks inevitably occur. In this unsecure environment, Intrusion Detection System (IDS) has become an essential component of computer security. It detects attacks with the aim of preserving systems from widespread damages and identifying vulnerabilities of the intruded system. There are two main categories of Intrusion Detection Systems: misuse detection and anomaly detection. Misuse detection systems represent attack in the form of pattern or signature while anomaly detection system differentiates normal traffic and attack. To classify network traffic in two classes various soft computing techniques are used [2][4][5][7][13][14]. Z. Muda [2] has used K-Means and Naive Bayes Learning Approach for Intrusion Detection. Fatin [3] proposed data mining techniques to detect intrusion. Ghanshyam [4] and Shailendra [7] used support vector machine for intrusion detection. While many other researchers [8][9][13][14][15][16] proposed Genetic Algorithm based intrusion detection system.

2. GENETIC ALGORITHM BASED INTRUSION DETECTION

S. Selvakani [8] proposed Intrusion Detection System based on Genetic Algorithm. For evolving and testing new rules for intrusion detection system

they have used KDD99 Cup training and testing dataset. In their approach Genetic algorithm was used to obtain classification rules for intrusion detection while correlation technique was used to identify the most important features of network connections.

Kunjal [9] design and develop a system to automatically evolve rules through genetic-fuzzy approach. Their work highlights the advantages of genetic and fuzzy hybridization and proposes a framework to evolve rules. The application is an intelligent system design to identify students' different skills in education domain. The architecture of evolving rule based model using genetic-fuzzy approach can also be applied to various domains like advisory systems, decision support systems, data mining systems, and control and monitoring systems, etc.

Ahmed [13] proposes optimization using Genetic Algorithms for the Security Audit Trail Analysis Problem, which was proposed by L. Mé in 1995 and improved by Pedro A. Diaz-Gomez and Dean F. Hougen in 2005. To classify attacks in "Certainly not existing attacks class", "Certainly existing attacks class" and "Uncertainly existing attacks class". The proposed idea is to divide the 3rd class to independent sub-problems easier to solve.

M. Sadiq [14] demonstrated that Genetic Algorithm can be effectively used for formulation of decision rules. The attacks which are more common can be detected more accurately. Rule based classification of DoS and Probe attacks can be used for effective monitoring of the network. Application of GA as compare to expert based knowledge is more fruitful as different possible combination of attribute is tested against training data and later validated through test data.

3. CHALLENGES

Genetic Algorithm has proven its potential in last decade and become the powerful weapon of many researchers. It is an evolutionary algorithm based on Darvin's theory. Genetic Algorithm becomes proven technology in optimization, searching, classification, and many more areas. This simple and powerful technique raises several challenges when it comes to its efficient implementation. Major challenges in using Genetic Algorithm for Intrusion Detection are as under:

The first challenge is designing chromosome. Genetic Algorithm requires problem to be represented in the form of chromosome. Whether to represent chromosome as sequence of character (string) of binary format? Performance of

Genetic Algorithm heavily depends on this. Second challenge is designing mutation and crossover. How to design crossover? Whether to use single point crossover, two point crossover, uniform crossover, or roulette crossover? Similar questions are there with respect to mutation probability and population size. Too low or too high population, deteriorate the overall performance and sometime behaves worst.

Next challenge in this list is to design effective fitness function. Genetic Algorithm required feedback to accomplice its task. After each iteration, Genetic Algorithm requires feedback about current population. In current population which chromosomes are poor and need to replace with new breed of strong chromosome. Without efficient fitness function Genetic Algorithm based intrusion detection performs poor. Another challenge is number of iteration or when to stop? Genetic Algorithm is a heuristic algorithm. It returns good result rather than the best one. What will be the threshold value? With what certainty declare network traffic as attack?

Another challenge faced by intrusion detection researchers is testing their framework or methodology. Many [8][9][13][14] used KDD 99 Cup Dataset for evaluating Genetic Algorithm based intrusion detection system. In literature few criticized KDD 99 CupDataset as outdated. Yet another problem is real-time attack detection in heavy network traffic. Genetic Algorithm based intrusion detection should be accurate and fast enough to detect attack in real-time. Complex algorithm will increase accuracy but demands more computation time. This will result in failure to detect attack in real-time. Simple and faster algorithm will work real-time but with less accuracy. Mobile and wireless networks have raise security concerns and open up new opportunity for intrusion detection.

4. SAMPLE IMPLEMENTATION

We have designed a simple Genetic Algorithm based solution for intrusion detection. It consists of two parts. In first part train the Genetic Algorithm based intrusion detection system with training data. In second part measure accuracy of algorithm by testing it using test dataset. Sample code is given below:

```
i = 0 // i = number of iterations
initializeP(i) // P(i): population for iteration i
evaluatef(P(i)) // f(P(i)): fitness function
```

```
while(not termination condition)
{
        i = i+ 1
        select 2 parents p1 and p2 from P(i)
        perform crossover operation
        perform mutation operation to generate three offspring: nos1, nos2, and
        nos3

        // reproduce a new P(i)
        if(random number <pa)// pa: probability of acceptance
        {
                The offspring among nos1, nos2, and nos3 with the largest
                fitness value replaces the least fit member of the population
        }
        else
        {
        if(f(nos1) >smallest fitness value in the P(i))
                nos1 replaces the least fit member of the population
        if(f(nos2) >smallest fitness value in the updated P(i))
                nos2 replaces the least fit member of the population
        if(f(nos3) >smallest fitness value in the updated P(i))
                nos3 replaces the least fit member of the population

        evaluate f(P(i))
        }
}
```

A simple fitness function is used to guide Genetic Algorithm.

$$f(x) = \alpha/A - \beta/B \qquad (1)$$

where α is the number of correctly detected attacks, A is the total number of attacks, β is the number of false positives, and B is the total number of normal connections. The fitness function value range was over the closed interval [-1, 1] with -1 being the weakest possible fitness and 1 being the best. A high correct detection rate and a low false positive rate results in a high score on the fitness function. Low detection rate or high false positive rate returns low scores on the fitness function.

5. RESULTS

The KDD Cup 99 dataset is popularly used as a benchmark dataset in several different researches [3] [5] [7] [8]. The KDD Cup 99 intrusion detection Data Set,

which is based on DARPA 98 Data Set, provides labeled data for researchers working in the field of intrusion detection and is the only labeled dataset publicly available. This Data Set is a benchmark Data Set for comparing performance of various IDSs [2][3][5][7][8]. For analyzing GA based intrusion detection, we have also used this benchmark Data Set. Following is the Class Distributions of 10% KDD99 Data Set:

Table-1 Class Distributions of 10% KDD99	
Class	Number of Connections
Normal	97277
DoS	391458
U2R	52
R2L	1126
Probe	4107
Total	494021

In our experiment we have tested our sample Genetic Algorithm based solution. We have focused on detection rate which is computed as the ratio between the number of correctly detected attacks and the total number of attacks.

Table-2 Experimental Results	
Class	Detection Rate
Normal	96.22 %
DoS	97.46 %
U2R	48.54 %
R2L	39.85 %
Probe	97.57 %

These experimental results indicate that Genetic Algorithm is a potential algorithm for intrusion detection. Genetic Algorithm performed well for Normal, DoS, and Probe attack class while performing poor for U2R and R2L attack classes. Possible reason for poor performance with U2R and R2L is insufficient training data. These results also indicate Genetic Algorithm has good learning capability which is required for indentifying new attacks quickly.

6. CONCLUSION

The study shown that Genetic Algorithm can be effectively used for intrusion detection. The attacks which are more common can be detected more accurately. It is also conveyed that implementation of algorithm contributes in higher accuracy. In this paper we surveyed various Genetic Algorithm based implementations for intrusion detection. Various challenges for Genetic

Algorithm based intrusion detection are also discussed. We have developed a simple Genetic Algorithm based intrusion detection system. We have used KDD cup 99 dataset for evaluating the performance of the simple system and experimentation results showed that the proposed system is having potential to detect various intrusions in computer networks.

REFERENCES:

1. Sumit A. Khandelwal, Shoba. A. Ade, Amol A. Bhosle and Radha S. Shirbhate, "A Simplified Approach to Identify Intrusion in Network with Anti Attacking Using .net Tool", International Journal of Computer and Electrical Engineering, Vol. 3, No. 3, June 2011
2. Z. Muda, W. Yassin, M. N. Sulaiman, and N. I. Udzir, "A K-Means and Naive Bayes Learning Approach for Better Intrusion Detection", Information Technology Journal 10(3): 648-655, 2011
3. 3Fatin Norsyafawati Mohd Sabri, Norita Md.Norwawi, and Kamaruzzaman Seman, "Identifying False Alarm Rates for Intrusion Detection System with Data Mining", IJCSNS International Journal of Computer Science and Network Security, VOL.11 No.4, April 2011
4. Ghanshyam Prasad Dubey, Prof. Neetesh Gupta, Rakesh K Bhujade, "A Novel Approach to Intrusion Detection System using Rough Set Theory and Incremental SVM", International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-1, Issue-1, March 2011
5. Ritu Ranjani Singh, Neetesh Gupta, Shiv Kumar, "To Reduce the False Alarm in Intrusion Detection System using self Organizing Map", International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-1, Issue-2, May 2011
6. P.Rajapandian, Dr.K.Alagarsamy, "Intrusion Detection in Dos Attacks", International Journal of Computer Applications (0975 – 8887) Volume 15– No.8, February 2011
7. Shailendra Kumar Shrivastava, Preeti Jain, "Effective Anomaly based Intrusion Detection using Rough Set Theory and Support Vector Machine", International Journal of Computer Applications (0975 – 8887) Volume 18– No.3, March 2011
8. S. Selvakani Kandeeban, R. S. Rajesh, "A Mutual Construction for IDS Using GA", International Journal of Advanced Science and Technology Vol. 29, April, 2011
9. Kunjal Mankad, Priti Srinivas Sajja, and Rajendra Akerkar, "EVOLVING RULES USING GENETIC FUZZY APPROACH - AN EDUCATIONAL CASE STUDY", International Journal on Soft Computing (IJSC), Vol.2, No.1, February 2011
10. VEGARD ENGEN, "MACHINE LEARNING FOR NETWORK BASED INTRUSION DETECTION", PhD thesis, Bournemouth University, June 2010
11. R. Shanmugavadivu, Dr.N.Nagarajan, "NETWORK INTRUSION DETECTION SYSTEM USING FUZZY LOGIC", Indian Journal of Computer Science and Engineering (IJCSE)
12. S.Sethuramalingam, Dr.E.R. Naganathan, "HYBRID FEATRUE SELECTION FOR NETWORK INTRUSION", International Journal on Computer Science and Engineering (IJCSE)
13. Ahmed AHMIM, Nacira GHOUALMI, Noujoud KAHYA, "Improved Off-Line Intrusion Detection Using A Genetic Algorithm And RMI", (IJACSA) International

Journal of Advanced Computer Science and Applications, Vol. 2, No.1, January 2011
14. M. Sadiq Ali Khan, "Rule based Network Intrusion Detection using Genetic Algorithm", International Journal of Computer Applications (0975 – 8887) Volume 18– No.8, March 2011
15. Ian Stewart, "A Modified Genetic Algorithm and Switch-Based Neural Network Model Applied to Misuse-Based Intrusion Detection", Master of Science Thesis, Queen's University, Kingston, Ontario, Canada, February 2009
16. Zorana Bankovic,"A Genetic Algorithm-based Solution for Intrusion Detection", Journal of Information Assurance and Security, (2009) 192-199
17. Ian Stewart, "A Modified Genetic Algorithm and Switch-Based Neural Network Model Applied to Misuse-Based Intrusion Detection", MS Thesis, Queen's University, Kingston, Ontario, Canada, February 2009
18. S. SELVAKANI and R.S.RAJESH, "Escalate Intrusion Detection using GA - NN", Int. J. Open Problems Compt. Math., Vol. 2, No. 2, June 2009
19. Zorana Bankovic, José M. Moya, Álvaro Araujo, Slobodan Bojanic and Octavio Nieto-Taladriz, "A Genetic Algorithm-based Solution for Intrusion Detection", Journal of Information Assurance and Security 4 (2009) 192-199
20. H. Güneş Kayacık, A. Nur Zincir-Heywood, Malcolm I. Heywood, "Selecting Features for Intrusion Detection: A Feature Relevance Analysis on KDD 99 Intrusion Detection Datasets", the NIMS Laboratory, http://www.cs.dal.ca/projectx, 2006
21. Russell Meyer, “Challenges of Managing an Intrusion Detection System (IDS) in the Enterprise”, As part of Information Security Reading Room, SANS Institute, http://www.sans.org, June 2011
22. Valerie Vogel, "Network and Host Security Implementation". Retrieved June 2011 from https://wiki.internet2.edu/confluence/display/secguide/Network+and+Host+Security+Implementation+(Stage+1)

AUTHORS’ PROFILE

Prof. Dharmendra G. Bhatti is an Associate Professor of MCA program at Shrimad Rajchandra Institute of Management and Computer Application, Bardoli, Gujarat, India. He received his MCA degree from South Gujarat University in the year 2000. His research area includes Network Security, Soft Computing Techniques, and Network Administration. He is having 11 years of experience and pursuing Ph.D. in Intrusion Detection using Soft Computing. He has published 1 research papers in International journal and 2 research papers in National journal. He has presented 7 research papers in National conferences/seminars and attended 9 winter/summer school. He is life member of Computer Society of India and achieved 6 IBM Certifications. He manages wired/wireless network of more than 1000 computers. He also handles windows active directory infrastructure, database servers, web server, proxy server, email service, and software licensing.

Dr. P. V. Virparia - Refer page number 86 for author profile.

Dr. Bankim Patel (Ph. D.) is a Director at Shrimad Rajchandra Institute of Management and Computer Application, Bardoli, Gujarat, India. His interest area includes Intelligent Information Systems. He is having 20 years of academic and 17 years research experience. He has published 7 research papers in International journal and 31 research papers in National journal. He has presented 1 research papers in National conferences/seminars. He has published 2 articles and 3 books. He has gained 8 honors and awards.