Efficient Algorithm for Feature Intruder Detection System

Vahid Kaviani

[1]*Faculty of Computer Engineering, Najafabad Branch, Islamic Azad University, Esfahan, IRAN*

---***---

Abstract - *This paper was carried out to use the Gravitational Search Algorithm for feature selection in IDS to selectively choose significant features which represents categories of network such as DoS, Probe, U2R and R2L and to improve the accuracy and effectiveness of feature selection and to have better detection. This paper aimed to study trends of feature selection in IDS and to implement BGSA for selectively choose features for IDS and to test and validate the performance and feedback of BGSA. In this paper, it validates and evaluates the BGSA algorithm and focuses on the feature selection by implementing of BGSA. The results of BGSA program proves and shows that the selected features which proposed by BGSA in terms of accuracy and efficiency are quite acceptable.*

***Key Words*: Denial-Of-Service (DOS), Intrusion Detection System (IDS), Detection Algorithm, Security, Privacy, Attacks.**

1. INTRODUCTION

The progress of computer technology has affected communication technology. From 1980s, many devices have been invented and developed. The progress in the network technology changes the way of communication and data distribution in the world because many businesses and companies use this technology for trading and marketing their products and contacting their partner and customers properly. Due to the completion and surviving in this generation among all organizations, the importance of safeguard and other countermeasures to stop penetration of intruders to their sensitive or critical information has been raising significantly [1-6]. To begin with definition in terms of attack, intruder is somebody who can maliciously interrupt, captures, modify, steal or delete important information in the computers and applications by network access or by direct access like run executable code in PC [7-9]. Attackers use different resources of victim to do the attack. Specifically, they misuse hardware vulnerabilities or software weakness to penetrate the system [10-13].

This paper was carried out to use the Gravitational Search Algorithm for feature selection in IDS to selectively choose significant features which represents categories of network such as DoS, Probe, U2R and R2L and to improve the accuracy and effectiveness of feature selection and to have better detection[14]. This paper aimed to study trends of feature selection in IDS and to implement BGSA for selectively choose features for IDS and to test and validate the performance and feedback of BGSA. In this paper, it validates and evaluates the BGSA algorithm and focuses on the feature selection by implementing of BGSA [15, 16]. The results of BGSA program proves and shows that the selected features which proposed by BGSA in terms of accuracy and efficiency are quite acceptable [15]. Also, they IDS system in this model is useful to apply for existing wireless technology to detect different attack with variety of wireless standards and technologies such as WSN, MANET, BANET, BYOD and physical system [17-20].

1.1 Types of Intrusion Detection Systems by Detection Techniques

Host-based IDS: Based on the sources of data, intrusion detection systems can be divided into two major classes, host-based and network-based. In the first kind of systems, the intrusion detection mechanism is installed on the local host/terminal. By examining the status of audit information on system's behaviour, the system finds signs of intrusion and can then protect its own local machine. The audit information can be obtained from different sources such as system logs and activities, application logs, and target monitoring [21, 22].

These logs could be UNIX logs, NT/2000/XP logs, firewall logs, router logs, web server logs, and FTP logs. The intrusions can be critical file modifications, segmentation fault errors recorded in logs, crashed services or extensive usage of the processors [23]. From the system point of view, all users are considered as local clients to the target environment.

Network-based IDS: In the network base is not just the host-based intrusion detection system to protect its own host machine by examining audit trail, network-based intrusion detection system protects the entire environment of the network by monitoring all the activities from both incoming and outgoing packets of the network. By analysing the traffic data that goes through the network, the potential and possible intrusions can be identified [10]. In general, the network traffic that needs to be monitored is quite heavy and large even in small networks. With good location for sensors on the network, instead of central sensor in the network, deploying sensors in different locations to achieve better efficiency is more effective [24].

Network-based IDS monitors any number of hosts on a network by inspecting the audit trails of multiple hosts [23]. Since attempted intrusions can happen via the network, network-based IDS needs to monitor multiple events generated on several hosts to integrate sufficient evidence. Since most of the hosts are networked and attacks can also be launched from remote, this study focuses on network-based IDS. Both host-based and network based IDSs mainly employ two detection techniques; anomaly and misuse.

2. Problem Situation and Solution Concept

It is very difficult to keep the model and signatures updated in all the time. So far this question will come to in mind that how and which approach will cause the best optimization for retraining and updating the IDS which deals with large traffic volume and new pattern of attacks and also difficulty with differentiation of normal and abnormal behaviour. Based on the other researches done in this area, it is clear that the effectiveness of an IDS model relies on retraining of the reference models and enhancing the recognition of classifiers. One of most important issue in IDS, in order to have better detection, is Feature selection.

Table -1: Summary of Problem Situations and Solution Concepts

Problem situation	**Solution Concept**
High overhead due to enormous data which Have to inspect by IDS. Some of them are unnecessary and redundant pattern.	Selective recognition through traffic filtering To filter out typical normal http connection. Improve data representation through feature selection by B-GSA.
loss of accuracy and time consumption in detection	Using B-GSA in feature selection of Intrusion Detection System

3. Classifying Features SVM

Support Vector Machine (SVM) is a learning method based on the Structural Risk Minimization principle from statistical learning theory. The principle idea of an SVM is to separate classes or attacks with a surface that maximizes the margins between them. It is a powerful classification learning approach which applies the following concept: non-linear input vectors are mapped through a very high dimension feature space where the linear decision of the input vectors is computed in this feature space. By dividing the high-dimensional space into different boundaries or subspaces, SVM maximizes the classification according to the generalized boundary. The SVM will extract and differentiate each feature of attacks and train and test datasets.

4. Evaluation of Feature Selection Algorithm

In order to test the effectiveness of feature selection, it is necessary to evaluate the fitness of each feature. If the fitness value is less than the present fitness value, repeat Step 1 to reach proper value otherwise the feature was correctly selected.

5. Parameters in System Model

Table -2: Summary of parameters used in this model

Variable	Description	Random, calculate by V,a
V	Velocity of agents	Non ,Calculate by M,X,R
a	Acceleration of masses or agents	Non, Calculate by Fitness of Subsets
M	Mass calculates by position of agents	41 is constant
m	Dimension is equal to the number of features in binary search space	50 is constant
N	Number of agents.	Random, calculate by V,a
X	Number of agents.	1
R	Distance between agents in search space.	Non ,range from 1 to 41
Fbest	The best Values of fitness' Agent	Non ,range from 1 to 41
GFbest	selected features in search space over iterations	non
BestChart	The best Values of fitness' Agent so far Chart over iterations	non
SubFitness Chart	Fitness of All Agents in all iterations	non
TFitnessC hart	Overall accuracy of dataset	Non

6. Results for Normal Class

Table 3 shows the best agents' values in iteration 1, 2, 4, 5, 9 and 17 which satisfy the first criterion which is the comparison of the best fitness subsets' values in each iteration. It can be observed from the table that the fitness value is increases from one iteration to another. The point is that in some iteration such as 3,6,7,8,10,11,12,13,14,15 and 16, the fitness values weren't satisfied the criterion. It means the best subsets, which extracted among 850 attempts in figure. In another word 850 times fitness function called through 17 iteration * 50 agents in each round in order to find best subsets. Therefore in order to make new proposed features, program respectively got 61, 65, 68, 70, 72 and 74 percentage accuracy in fitness process. The results of the accuracy rate and purposed features in all five respective network traffics (Normal, Probe, DoS, U2R and R2L) were produced by BGSA program for feature selection to minimize the recognition time and to omit redundant and high overload features. The result demonstrated in figure 1 and 2.

Table 3. Best Fitness Values in Class 1

Iteration	Best Fitness of Agent's Value
1	61.46343
2	65.85363
4	68.04882
5	70.24392
9	72.43902
17	74.63412

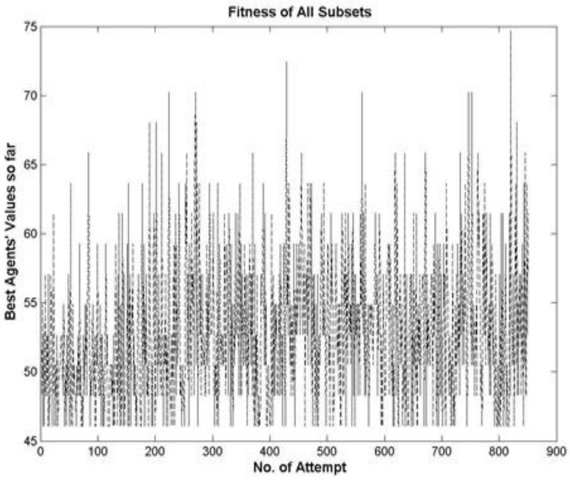

Figure 1. All Fitness Values in Class1

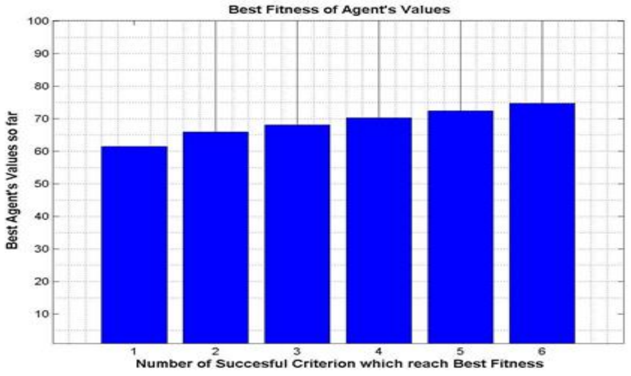

Figure 2. Best Fitness Values in Class 1

3. CONCLUSIONS

This paper was carried out to use the Gravitational Search Algorithm for feature selection in IDS to selectively choose significant features which represents categories of network such as DoS, Probe, U2R and R2L and to improve the accuracy and effectiveness of feature selection and to have better detection. The program successfully minimized the time for recognition process which is using the concept of selective recognition and minimal feature set. In the most high-dimensional issues, Selection of the influence features and remove of other features can greatly raise the accuracy of classification and exactly reduce the complexity of data processing at different stages.

REFERENCES

[1] W.-C. Lin, S.-W. Ke, and C.-F. Tsai, "CANN: An intrusion detection system based on combining cluster centers and nearest neighbors," *Knowledge-based systems,* vol. 78, pp. 13-21, 2015.

[2] P. S. Kenkre, A. Pai, and L. Colaco, "Real time intrusion detection and prevention system," in *Proceedings of the 3rd International Conference on Frontiers of Intelligent Computing: Theory and Applications (FICTA) 2014*, 2015, pp. 405-411.

[3] P. R. K. Varma, V. V. Kumari, and S. S. Kumar, "Feature Selection Using Relative Fuzzy Entropy and Ant Colony Optimization Applied to Real-time Intrusion Detection System," *Procedia Computer Science,* vol. 85, pp. 503-510, 2016.

[4] G. Gu, P. A. Porras, and M. Fong, "Method and apparatus for detecting malware infection," ed: Google Patents, 2015.

[5] A.-S. K. Pathan, *Security of self-organizing networks: MANET, WSN, WMN, VANET*: CRC press, 2016.

[6] M. Razzaque, A. Salehi, and S. M. Cheraghi, "Security and privacy in vehicular ad-hoc networks: survey and the road ahead," in *Wireless Networks and Security*, ed: Springer, 2013, pp. 107-132.

[7] D. C. Harrison, W. K. Seah, and R. Rayudu, "Rare Event Detection and Propagation in Wireless Sensor Networks," *ACM Computing Surveys (CSUR),* vol. 48, p. 58, 2016.

[8] S. K. Roy, A. Roy, S. Misra, N. S. Raghuwanshi, and M. S. Obaidat, "AID: A Prototype for Agricultural Intrusion Detection Using Wireless Sensor Network," in *2015 IEEE International Conference on Communications (ICC)*, 2015, pp. 7059-7064.

[9] A. Salehi Shahraki, M. Razzaque, P. Naraei, and A. Farrokhtala, "Security in wireless sensor networks: issues and challenges," 201

[10] E. Eskin, A. Arnold, M. Prerau, L. Portnoy, and S. Stolfo, "A geometric framework for unsupervised anomaly detection," in *Applications of data mining in computer security*, ed: Springer, 2002, pp. 77-101.

[11] H. Hoffmann, M. J. Daily, G. D. Holland, and K. El Defrawy, "System and method for deep packet inspection and intrusion detection," ed: Google Patents, 2016.

[12] S. A. Salehi, M. Razzaque, P. Naraei, and A. Farrokhtala, "Security in wireless sensor networks: Issues and challanges," in *Space Science and Communication (IconSpace), 2013 IEEE International Conference on*, 2013, pp. 356-360.

[13] A. Salehi , Sayyadi, Hamed, Hussein Amri, Mohamad, & Nikmaram, Mehrnaz, "Survey: video forensic tools" *Journal of Theoretical and Applied Information Technology,* vol. 47, 2013.

[14] A. Sal and M. NIKMARAM, "Human errors in computer related abuses" *Journal of Theoretical and Applied Information Technology,* vol. 47, 2013.

[15] M. S. Hoque, M. Mukit, M. Bikas, and A. Naser, "An implementation of intrusion detection system using genetic algorithm," *arXiv preprint arXiv:1204.1336,* 2012.

[16] W. Lee, S. J. Stolfo, P. K. Chan, E. Eskin, W. Fan, M. Miller, *et al.*, "Real time data mining-based intrusion detection," in *DARPA Information Survivability Conference &amp; Exposition II, 2001. DISCEX'01. Proceedings*, 2001, pp. 89-100.

[17] S. S. Ahmad, S. Camtepe, and D. Jayalath, "Understanding data flow and security requirements in wireless Body Area Networks for healthcare," in *2015 17th International Conference on E-health Networking, Application & Services (HealthCom)*, 2015, pp. 621-626.

[18] A. Salehi S, M. A. Razzaque, I. Tomeo-Reyes, and N. Hussain, "IEEE 802.15.6 Standard in Wireless Body Area Networks From a Healthcare Point of View," in *2016 Asia-Pacific Conference on Communications (APCC)*, Indonesia, 2016.

[19] S. Movassaghi, M. Abolhasan, J. Lipman, D. Smith, and A. Jamalipour, "Wireless body area networks: A survey," *IEEE Communications Surveys & Tutorials,* vol. 16, pp. 1658-1686, 2

[20] A. Salehi S, M. A. Razzaque, I. Tomeo-Reyes, and N. Hussain, "Efficient high-rate key management technique for wireless body area networks," in *2016 Asia-Pacific Conference on Communications (APCC)*, Indonesia, 2016.

[21] E. Rashedi, H. Nezamabadi-Pour, and S. Saryazdi, "GSA: a gravitational search algorithm," *Information sciences,* vol. 179, pp. 2232-2248, 2009.

[22] S. A. Salehi, M. Razzaque, P. Naraei, and A. Farrokhtala, "Detection of sinkhole attack in wireless sensor networks," in *Space Science and Communication (IconSpace), 2013 IEEE International Conference on*, 2013, pp. 361-365.

[23] S. Forrest, A. S. Perelson, L. Allen, and R. Cherukuri, "Self-nonself discrimination in a computer," in *IEEE Symposium on Security and Privacy*, 1994, pp. 202-212.

[24] Y. Li, J. Xia, S. Zhang, J. Yan, X. Ai, and K. Dai, "An efficient intrusion detection system based on support vector machines and gradually feature removal method," *Expert Systems with Applications,* vol. 39, pp. 424-430, 2012.