International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2015): 78.96 | Impact Factor (2015): 6.391

A Literature Survey on Existing Intrusion Detection Systems Based on Genetic Algorithm

Nagarapu Priyanka[1], Prikhil Agarwal[2]

[1, 2]Student, Department of Computer Science and Engineering, Institute of Technology, Guru Ghasidas Vishwavidhyalaya, (A Central University), Bilaspur ,Chhattisgarh ,India.

Abstract: ***With increase of millions of users on Internet day by day, it is very essential to maintain highly reliable and secured data communication between various corporations. Although there are various traditional security imparting techniques such as antivirus software, password protection, data encryption, biometrics and firewall etc. But still network security has become a main issue in various leading companies. So IDSs have become a essential component in terms of security, as it can detect various network attacks and respond quickly to such occurrences. IDSs are used to detect unauthorized access to a computer system. This paper describes various intrusion detection techniques using GA approach.The intrusion detection problem has become a challenging task due to the conception of miscellaneous computer networks under various vulnerabilities. Thus the damage caused to various organizations by malicious intrusions can be mitigated and even be deterred by using this powerful tool.***

Keywords: Genetic Algorithm (GA), Intrusion Detection System (IDS), Dataset, Network Security

1. Introduction

Intrusion detection systems (IDSs) are also called as protection systems. The key role of IDSs is to detect and log breaches and send a warning or alert the administrator or report the breach to central repository called "Security Information and Event Management (SIEM)"System. This system combines alerts from various tools to better distinguish malign activities from bad alarms .A copy of live traffic is sent to IDS from network tap to perform complex analysis and investigations and this traffic is not routed back again to the trusted network. This can also be called as passive monitoring because it does not operation live traffic. While the firewall monitors the incoming and outgoing traffic to and from an organization, some IDS can also be designed to identify internal attacks. Thus an IDS can be deployed at any strategic point in the network as per our requirements. Many services of computer technology have resulted in the increase of computer related threats. There is a difficulty to develop a complete solution for the security of attacks on complex systems, and attackers are trying to find new ways to ignore these interruption mechanisms. IDSs have become an unavoidable component of prevention systems. IDSs are built based on various detection and prevention techniques. Initially, systems mostly relied on signature-based identification where past attacks were analyzed and identified based on their characteristics..Later the limitation of detecting new attacks has overcome by using statistical anomaly detection techniques where the system could produce alerts based on the traffic criteria .IDSs are of two types: Anomaly intrusion detection and Misuse intrusion detection. Anomaly intrusion detection checks if there are any deviations from users normal behavior whereas in misuse intrusion detection patterns of known attacks are described and these patterns are used to identify the attacks.

The Intrusion Detection Systems can also be defined into two categories:- Host-based Intrusion Detection System and Network-based Intrusion Detection System. Host-based IDS evaluate the activities found on a single or multiple hosts whereas network-based IDS evaluate the activities captured from network communication. An Intrusion Detection System is a software application or device that observes network and system activities for vicious activities and outcomes the reports. The remaining paper is formed as follows: Section 2 describes problems with existing system. Section 3 gives the information of related works. Section 4 describes some dataset. Section 5 and 6 introduces the algorithm and aggregate result of related work. Section 7 describe experimental result part and section 8 is a conclusion part.

2. Problem With Existing System

IDSs deal with so many existing system such as snort, OSSEC, OSSIM, suricata, Bro, BASE, sguil, fragroute. These existing system undergoes at least two problems-

1) Audit trails acquire the information which is used by IDS. Data pass through a longer path from its source to the IDS and in the process attacker can destroy or modify the data. Furthermore, the IDS assume the performance of the system from the data collected, which can result in misplaced or misconception events. This is called as the fidelity problem.
2) The IDS moderately uses extra resources in the system it is checking even when there are no intrusions happening, because the parts of the IDSs have to be running all the time. This is called as resource usage problem.
3) The basics of the IDSs are resolved as separate programs, they are affected to moderate. An intruder can probably modify or disable the programs, analysis the IDS uncertain or useless. This is called as reliability problem.

3. Related Works

Melani J Middlemiss et al. (2003) [1] have completed the feature distillation with particular application to find intruder data using genetic algorithm. They have resolved GA which progresses weights for the characteristic of data set.

Ren Hui Gong et al. (2005) [2] have implemented GA to develop a set of rules and the support confidence framework is applied as fitness function to check the feature of each rule. The developed set of classification rules are then used to discover or categorize network intrusions in a real time environment. Jiu-Ling Zhao et al. (2005) [3] have discussed the computer network intrusion detection problem using clustering genetic algorithms. This algorithm can not only cluster the cases necessarily, but also discover the intruder.

Tao Xia et al. (2005) [4] discussed a hybrid method to detect the network attacks .This method is based on GA and information theory. Refining the traffic data to diminish the complexity, information theory is used.

Chi Hoon Lee et al. (2006) [5] discusses the novel feature selection method that inflates class separation between attack and normal patterns of network connections. They have fixated on choosing a robust feature subset for recover intruder. Saqib Ashfaq et al. (2006) [6] have implemented GA in order to generate set of rules for misuse detection in IDS. They have implemented a genetic algorithm to recognize these features.

Nalini N. and Raghavendra Rao G. (2006) [7] has discussed a method of Intrusion detection system using GA and GP. This method used to discover the class of intrusion. In this paper, they analyzed with PCA to deduct the number of features of a TCP connection.

Hua Zhou et al. (2007) [8] have discussed to increase the detection rate by using SVM and GA. They used genetic algorithm for optimization and feature selection and to detect intrusions they used SVM model. Yong Wang et al. (2009) [9] mainly deal with a fitness function, an efficient rule generator for denial of service attack. He did his implementation using GA toolbox granted by MATLAB (R14). He developed the GA using 4 m-files. The generating rules are acceptable for gradually changing misuse detection. S.N.Pawar and R.S.Bichkar (2012) [10] have introduced to use of GA with enumeration technique. We use enumeration technique while generating random population. This technique is very useful to resolve the identity of each gene. This technique considerably diminishes the computational time required for population generation.

Table 1: The GA approaches and GA parameters used by different researchers for intrusion detection

S.No.	Name of the Researcher and year	Approach used	GA parameters					% Detection
			Selection	Crossover	Mutation	Population	Generations	
1	Melani J Middlemiss et al. (2003)	Simple GA with A KNN classifier	Linear Ranking	0.6	0.0075	100	100	NA
2	Ren Hui Gong et al. (2005)	GA	Fitness Proportionate	0.5	0.02	500	5000	79.8
3	Jiu-Ling Zhao et al. (2005)	Clustering GA	NA	0.75	0.001	120	200	95
4	Tao Xia et al. (2005)	Hybrid method based On information theory and GA	NA	NA	0.01	1000	NA	99.2
5	Chi Hoon Lee et al. (2006)	GA	NA	0.6	0.05	30	50	62.9
6	SaqibAshfaq et al. (2006)	GA	Fitness Proportionate	NA	NA	200	50	96.4
7	Nalini N. and Raghavendra Rao G (2006)	GA and PCA	NA	NA	NA	NA	NA	93.1
8	Hua Zhou, XiangruMeng, Li Zhang (2007)	GA and SVM	Fitness Proportionate	0.8	0.05	100	300	98.9
9	Yong Wang et al. (2009)	GA	NA	NA	NA	NA	NA	95.4
10	S.N.Pawar And R.S.Bichkar (2012)	GA with enumeration	Fitness Proportionate	0.5	0.01	300	2000	98

Table 2: DataSet used by different researchers

S. No.	Name Of The Researcher And Year	DataSet
1	Melani J Middlemiss et al. (2003)	KDD CUP 99
2	Ren Hui Gong et al. (2005)	DARPA 1998
3	Jiu-Ling Zhao et al. (2005)	KDD CUP 99
4	Tao Xia et al. (2005)	KDD CUP 99
5	Chi Hoon Lee et al. (2006)	KDD CUP 99
6	Saqib Ashfaq et al. (2006)	KDD CUP 99
7	Nalini N. and Raghavendra Rao G (2006)	KDD CUP 99
8	Hua Zhou, Xiangru Meng, Li Zhang (2007)	KDD CUP 99
9	Yong Wang et al. (2009)	KDD CUP 99
10	S.N.Pawar and R.S.Bichkar (2012)	DARPA 1998

4. Dataset Used

4.1 DARPA 1998

DARPA 1998 Intrusion Detection Evaluation has been divided into two parts: an off-line evaluation and a real-time evaluation. All the researches have tested IDSs in the off-line evaluation using genetic algorithm.MIT Lincoln Laboratory has possessed and delivered the first standard data for the development of computer NIDS. The systems handled this data in batch mode and pursued to recognize attack sessions in the core of normal enterprises. DARPA1998 exists of tcpdump and BSM list files. In a list files each line coincides to a separate session. Each session describes to an exclusive TCP/IP connection between two computers. In the list files the first nine columns gives the information which recognizes the TCP/IP connection.

4.2 KDD CUP 99

KDD CUP 99 data is the part of the data composed from the MIT Lincoln Labs 1998 DARPA Intrusion Detection Evaluation Program and is deliberated benchmark data for reviewing IDSs. The data is available in smaller or in a number of full data sets.

5. Aggregate Analysis of Related Work

Table 1 gives the details of percentage (%) detection acquired by using the various genetic algorithm based approaches. It is clear that Tao Xia et al. (2005) introduced hybrid method which is based on genetic algorithm and information theory achieved intensely good results (99.25%).

S.N.Pawar and R.S.Bichkar (2012) mainly deals with enumeration technique in a genetic algorithm which is based on rule generation have achieved good results (98.06%).

Hua Zhou et al. described genetic algorithm for SVM model and feature selection have achieved good detection rates (98.97%).

Overall we can say that if we use genetic algorithm approach and other soft computing technique, the intrusion detection results acquired using genetic algorithm approach are better or are comparable to the results acquired by using other soft computing techniques.

6. Algorithm

Major steps to detect using GA
Algorithm - Rule Set Generation using GA
Input-Size of population, Network Audit Data, set of chromosomes

Output- Set of rules
Step 1) Initialization of population
Step 2) Initialization Crossover rate and Mutation rate
Step 3) While {
Number of generations defined is not reached
Step 4) for {
Each precalculated chromosome in population
Step 5) Evaluate the fitness function of new individual
Step 6) Allow only those fitness function that fit the fitness criteria
Step 7) End for loop}
Step 8) Discard some chromosome having worse fitness
Step 9) Implement crossover for recreation of new rule by swapping some bits
Step 10) Implement mutation by throwing some bits
Step 11) End while loop}
Step 12) Go to step 5, until the classification rules are not created.

7. Experimental Result and Analysis

Our system gives the various results. In this, Middle part gives the detail of displayed events. Events box shows total number of events in a selected 24-hour period, where red line gives the details of IP-based events, blue line shows DNS-based events and yellow line performs URL-based events. Sources box gives the details of number of events per top sources, with total number of sources on top. Threats box and trails box gives the details in pie chart form.

Table 3: Detail of Threat box and Trail box

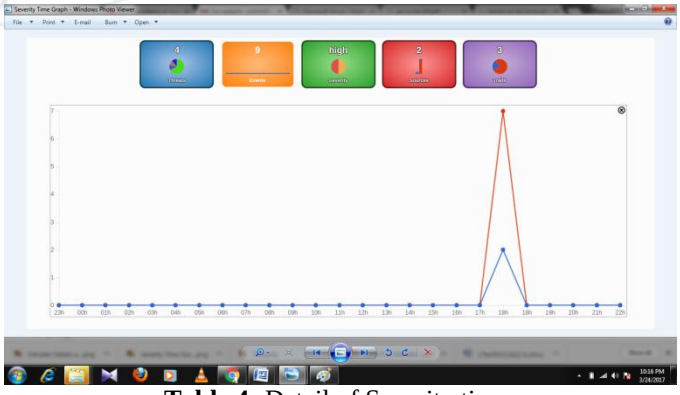

Table 4: Detail of Severity time

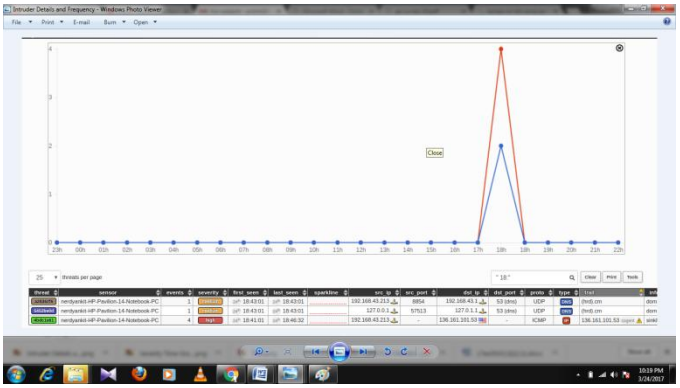

Table 5: Intruder Details

8. Conclusion

This survey an introduction to detect the intrusion using genetic algorithms. It targets on different type of IDS models. It is fulfilled that to implement the IDS we use various technique. This survey provides detailed information to the field of GA based intrusion detection. On the basis of survey GA is one of the productive techniques favorably in network intrusions detection.Our result also displayed the severity of each attack. It is resolved in intrusion detection system either for the suitable generating the set of rules or selecting activities.

References

[1] Melanie Middlemiss and Grant Dick, "Weighted Feature Extraction Using a Genetic Algorithm for Intrusion Detection", 2003 Congress on Evolutionary Computation (cec-03) 2003, pp.1669-1675.
[2] Ren Hui Gong, Mohammad Zulkernine and Purang Abolmaesumi, "A Software Implementation of a Genetic Algorithm Based Approach to Network Intrusion Detection", SNPD/SAWN'05, IEEE, 2005.
[3] Jiu-Ling Zhao, Jiu–Fen Zhao and Jian –Jun Li, "Intrusion Detection Based on Clustering Genetic Algorithm", International Conference Based on Machine Learning and Cybernetics, IEEE, Guangzhou, 2005, pp.3911-3914.
[4] Tao Xia, Guangzhi Qu, Salim Hariri and Mazin Yousif, "An efficient Network Intrusion Detection Method Based on Information Theory and Genetic Algorithm", IEEE, 2005.
[5] Chi Hoon Lee, Sung Woo Shin and Jin Wook Chung, "Network Intrusion Detection Through Genetic Feature Selection", SNPD, IEEE, 2006.
[6] Saqib Ashfaq, M.Umar Farooq and Asim Karim, "Efficient Rule Generation for Cost- Sensitive Misuse Detection Using Genetic Algorithms", IEEE, 2006.
[7] Nalini N and Raghavendra Rao G., "Network Intrusion Detection via a Hybrid of Genetic Algorithms and Principal Component Analysis", IEEE, 2006.
[8] Hua Zhou, Xingu Meng and Li Zhang, "Application of Support Vector Machine and Genetic Algorithm to Network Intrusion Detection", IEEE, 2007.
[9] Yong Wang, Dawu Gu, Xiuxia Tian and Jing Li, "Genetic Algorithm Rule Definition for Denial of Services Network Intrusion Detection", International Conference on Computational Intelligence and Natural Computing, IEEE, 2009, pp.99-102.
[10] S. N. Pawar and R. S. Bichkar, "Using Enumeration in a GA based Intrusion Detection", International Journal of Computer Applications (IJCA), October, 2012.