ISSN 2321-807X
Volume 12 Number17
Journal of Advances in chemistry

Intelligent Intrusion Detection System Using Genetic Algorithm

Dr.S.Kannan[1],A.Albert Martin Ruban[2],M.Arun[3]
Department of Information Technology, Kings College of Engineering,Thanjavur, Tamilnadu, India
Email::kannan.pot@gmail.com
Department of EEE, Kings College of Engineering,Thanjavur, Tamilnadu, India
Email: albertrubankings@gmail.com
Department of CSE, Kings College of Engineering,Thanjavur, Tamilnadu, India
Email:mailstoroyalarun@gmail.com

ABSTRACT

Intrusion detection is an essential and important technique in research field. One of the main challenges in the security system of large-scale high-speed networks is the detection of suspicious anomalies in network traffic patterns due to different kinds of network attack. We give attacks normally identified by intrusion detection systems. Differentiation can be done in existing intrusion detection methods and systems based on the underlying computational methods used. Intrusion detection methods started appearing in the last few years. In this paper we propose an Intrusion detection method using Genetic Algorithm (GA). In this research contribution of each of above mentioned techniques will be systematically summarized and compared that will allows us to clearly define existing research challenges, and to highlight promising new research directions.

Indexing terms/Keywords

Intrusion Detection System, Genetic algorithm.

Academic Discipline And Sub-Disciplines

Department of Information Technology,Department of Electrical and Electroniscs Engineering

SUBJECT CLASSIFICATION

Genetic Algorithm, Intrusion Detection system

TYPE (METHOD/APPROACH)

Genetic Algorithm, Intrusion Detection system experimental Research

1. INTRODUCTION

Intrusion is defined as an act of violating the confidentiality, integrity, or availability of a computer or a computer network system. Intrusion detection system (IDS) is one of most important system being used to detect the Internet attacks which can be either host based or network based. IDS plays a major role in maintaining and keeping information secure in any networking environment. An intrusion detection system records and monitors all inbound and outbound network traffic and identify suspicious activity which tries to break network security. Intrusion Detection System (IDS) can be software or a combination of software and hardware that automatically detects computer intrusions and reacts properly in order to protect computers and networks from activity occurring in your network [6]. Traditionally Intrusion detection approaches are classified into following categories.

- Misuse detection
- Anomaly detection
- Specification-based detection

1.1 Misuse detection: misuse based IDS uses information of already occurred attack patterns to identify attacks. Intrusions are detected by matching actual behavior recorded in audit trails with known suspicious patterns. So from its working it is clear that misuse detection is fully effective in detecting known attacks but it is useless when encountered with unknown or novel forms of attacks for which the signatures are not yet available.

Model consists of four major components: namely, Data collection, system profile, misuse detection and response. Data are collected from one or many data sources including audit trails, network traffic, system call trace, etc. Collected data are transferred into a format that is understandable by the other components of the system. The system profile is used to characterize normal and abnormal behaviors. The profiles are matched with actual system activities and reported as intrusions in case of deviations. Four classes of techniques are commonly used to implement misuse detection, namely pattern matching, rule-based techniques, state-based techniques, and data mining.

1.2 Anomaly Based: Anomaly based IDS uses normal instances as the base data to operate on. Any instance or behavior deviating from this normal behavior is termed anomalous and is categorized as an attack. Anomaly detection studies start by forming an opinion on what the normal attributes for the observed objects are, and then decide what kinds of activities should be flagged as intrusions and how to make such particular decisions.

***1*.3 Specification-based detection:** Specification based approach is different from misuse and anomaly based techniques. Instead of learning system behaviors in specification-based systems the experts' knowledge determines the operating limits of a system. Once the correct (or allowed) system behavior is specified. The events deviating from the specification would generate an alert [7].

2. LITERATURE SURVEY

The Intrusion Detection System has undergone rapid changes and is using new evolved techniques to generate better results. There are several approaches for solving intrusion detection problems assizes on the network.

Chittur at al. applied detection of computer intrusions and malicious computer behavior and analyzed the effectiveness of a Genetic Algorithm and compared their results with previous intrusion detection techniques [18].

Li described a method using GA to detect anomalous network intrusion. His approach includes both Quantitative and categorical features of network data for deriving classification rules. He addressed the Factors affecting to Genetic Algorithm are in detail. The inclusion of quantitative feature can increase detection rate but experimental results are not available [19].

Xia et al. detected anomalous network behaviors based on information theory by using GA. Some network features can be identified with network attacks based on mutual information between network features and type of intrusions and then using these features a linear structure rule and also a GA is derived. The advantage of the approach of using mutual information and resulting linear rule seems very effective because of the reduced complexity and higher detection rate. The approach has disadvantage that it considered only the discrete features [20].

Abdullah showed a GA based performance evaluation algorithm to network intrusion detection. The traffic data was filtered by information theory in his approach [21].

Lu and Traore used support-confidence framework as the fitness function and accurately classified several network intrusions. Disadvantage of their approach is that, use of genetic programming made the implementation procedure very difficult and also for training procedure more data and time is required [22].

Gong et al. presented an implementation of GA based approach to Network Intrusion Detection using GA and showed software implementation. In this approach he derived a set of classification rules using a support-confidence framework to judge fitness function [23].

T. Bharat et al. B. Uppalaiah, [24] presents the Genetic Algorithm for the Intrusion detection system for detecting DoS, R2L, U2R, Probe from DD99CUP data set. The architecture of the system along with implementation of the software for the proposed technique is also discussed. The time to get thorough with three features to describe the data will be reduced with a combination of Genetic Algorithm based IDSs. there system is more flexible for usage in different application areas with proper attack taxonomy. Genetic Algorithm detects the intrusion while correlation techniques identify the features of the network connections. The results shows that we have specified set of rules and high Dos, R2L, U2R, Probe attack detect rate. In optimizing the parameters present in the algorithm reduces the training time.

SaumyaChandra et al ,Srinivasa K G.[25] presents IGIDS, where the genetic algorithm is used for pruning best individuals in the rule set database. The process makes the decision faster as the search space of the resulting rule set is much compact when compared to the original data set. This makes IDS faster and intelligent.

Chetan Kumar and Anup Goyal [26] has presented a machine learning approach known as Genetic Algorithm (GA), to identify such harmful/attack type of connections. The algorithm takes into consideration different features in network connections such as type of protocol, network service on the destination and status of the connection to generate a classification rule set. Each rule in rule set identifies a particular attack type. For this experiment, they implemented a GA and trained it on the KDD Cup 99 data set to generate a rule set that can be applied to the IDS to identify and classify different types of attack connections.

Brian E. Lavender [27] proposed the integration of genetic algorithms (GA) into SNORT to enhance SNORT at performing Network Intrusion Detection (NID).

Shaik Akbar et al. [28] presents an algorithm which identifies damaging/attack type connections called Genetic Algorithm. The algorithm considers different features by protocol type, duration, src_bytes in generating a rule set. The Genetic Algorithm is trained on the KDDCUP99 Data Set in order to generate a collection of rules which applied on Intrusion Detection System identifies different types of attacks.

3. GENETIC ALGORITHM

"A Genetic Algorithm (GA) is a programming technique that mimics biological evolution as a problem-solving strategy. It is based on Darwinian's principle of evolution and survival of fittest to optimize a population of candidate solutions towards a predefined fitness. Genetic Algorithms are biologically inspired search heuristics that employs evolutionary algorithm

techniques like crossover, inheritance, mutation, selection etc. So, genetic algorithms are capable of deriving classification rule and selecting optimal parameters for detection process. GA 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 [29].

3.1 Parameters used in genetic algorithm

3.1.1 Fitness Function: The fitness function evaluates the quality of a particular solution. The fitness function is used to select the best solution among all the solutions in the population. The fitness function should be an optimized value.

3.1.2 Selection: Selection is the process of choosing solution with better fitness function than their counterparts. In the selection phase the solutions having better fitness function over other solutions are selected and the rest are discarded.

3.1.3 Crossover: Crossover is the phase in which two solutions exchange one of their characteristics with the other in the pair at a randomly selected crossover point, where the crossover probability is between 0.6 and 0.9. The solutions selected for crossover operation should be different.

3.1.4 Mutation: Mutation is a process by which some random bits in a solution are changed. This is done mainly to maintain the genetic diversity of the solutions [31].

When genetic algorithm is used for problem solving, three factors will have impact on the effectiveness of the algorithm, they are

- The selection of fitness function
- The representation of individuals and
- The values of the genetic parameters

Major Steps in Genetic Algorithm

Algorithm: Rule set generation using genetic algorithm.

Input: Number of generations, and population size.

Output: A set of classification rules

1. Initialize the population

2. Check the fitness function

3. Select only those rules that that meets the fitness criteria.

4. Perform crossover for reproduction of new rule by exchanging some bits

5. Perform mutation by flipping some bits

6.Again go to line 2, until the specified numbers of rules are not generated

Data representation in genetic algorithm: Genes should be represented in some format using different data types such as byte, integer and float. Also they may have different data ranges and other features, knowing that the genes are generated randomly, in each population generating iteration. The conditions to detect the intrusion is generally the current network traffic or connection details like source IP address, destination IP address, port numbers (like TCP, UDP), duration of the connection, protocols used. For genetic algorithm to work in IDS each of above mentioned condition is converted into chromosome. Each chromosome is evaluated by a fitness function to determine the solution's quality; better-fit solutions survive and produce offspring, while less-fit solutions are culled from the population. Genetic algorithms can be used to evolve rules for the network traffic; these rules are usually in the following form:

"If {condition} then {act}"

It basically contains if-then clause, a condition and an act. The conditions usually matches the current network behavior with the one stored in the in the IDS such as comparing an intruder source IP address and port number with one already stored in the system. The act could be an alarm indicating that the intruders IP and Port numbers are related to an attacker who is previously known in the system. Example if (duration = "0:0:1" and protocol = "finger" and source_ port = 18989 and destination port = 79 and source_ip = "99.19.99.19" and destination_ip = "192.168.254.10") then (attack name = "Neptune"). Above rule specifies that if a network packet is originated from IP address 172.16.98.10 and port number 1820 and send to IP address 97.4.11.36 at port number 80 using finger protocol for duration of connection 1 second then most likely it is Neptune attack which eventually make destination host out of service.

3.1 GENETIC ALGORITHM PROCESS

GA advances the number of inhabitants in chromosomes (people) as the procedure of characteristic selection [30]. It generate(s) new chromosome(s) (posterity) amid its procedure. GA process utilizes an arrangement of hereditary administrators (determination administrator, hybrid administrator and change administrator), and assess chromosome utilizing the wellness function.GA comprises of populace of chromosomes that imitated over arrangement of eras as indicated by their wellness in a domain. Chromosomes that are most fit are well on the way to survive, mate, and bear children.GA end the procedure by characterize altered maximal number of eras or as the accomplishment of a satisfactory wellness level, or if there are no upgrades in the populace for some settled eras, or for some other reason. The standard GA procedure is appeared in figure 1. It contains different steps which incorporate: encoding chromosomes, creating introductory populace, wellness capacity assessment, then applying one of the administrators. The procedure will stop when we get the best people

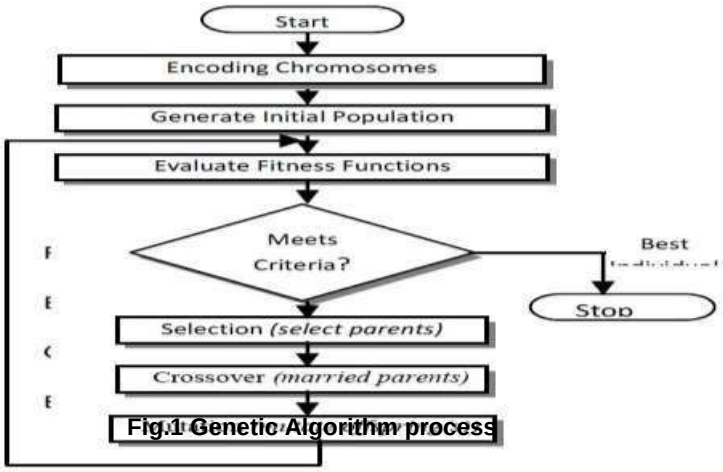

Fig.1 Genetic Algorithm process

4. EXPERIMENTAL SETUP

The IP locations are created on an extensive scale. In Intrusion Detection method the IP addresses making annoyance are should be recognized. The Genetic calculation utilized is to recognize the gatecrashers and the last rundown will be sent to firewall. Consequently the system will be shielded from assaults and will expand the execution of the system. The point of this paper is to distinguish the gatecrashers and create the rundown of IP locations with its event check. The Genetic calculation will distinguish the frail and solid interlopers and produce the last rundown and would piece them on firewall. The slightest access rights given to such clients would expand the execution of the system.

4.1 TOOLS

For this experiment we have used java as the frontend software which code overall GA operator and there evolution process. The training data is stored into the wamp server which is used as the backend to the system. For this experiment we used windows based Dell computer with dual core processor system having 120 GB hard disk space and 2 GB RAM to execute the computer program.

4.2 GENETIC OPERATORS AND PARAMETERS

Table 1: Parameter Setting For GA Algorithm

SETTING TYPE	VALUES
Population size	100
Evolution generation	30
Selection	Fitness-proportion
Crossover	One point
Mutation	Real number
Generations	30

5. RESULT AND PERFORMANCE

From the above experiment, we have able to create a rule base that could successfully categories harmful and harmless

connection types. We have shown the resultant figures below by applying 100 connection entries respectively to the proposed system. After that we were able to get around 99% of accuracy to classify the connections types.

Table-2: Proposed System For Filtration

Source -IP	Destination-Port	Source-Port	Destination-Port
172.16.98.10	97.4.11.36	1820	80
172.16.98.10	97.42.11.136	1820	80
172.16.98.10	97.43.10.126	1355	65
172.16.98.10	185.4.11.26	1300	35
172.16.98.10	97.40.11.36	1544	65
172.16.98.10	97.40.11.36	1544	85
172.16.98.10	233.255.255.250	68	1320
172.16.98.10	233.255.255.250	1032	64
172.16.98.10	197.14.11.36	68	1034

6. CONCLUSION

In this paper we have successfully evolved the rule set which can detect existing as well as new intrusions. So as the result generated; the system can be integrate with any of the IDS system to improve the efficiency and the performance. The system can also be able to integrate to the input to the firewall system. In this paper, we have discussed the GA processes and evolution operators also discussed the overall implementation of GA into proposed system. The various operators like selection, crossover and mutation is also discussed.

REFERENCES

1. Mouna Jouini , Latifa Ben Arfa Rabai , & Anis Ben Aissa (2014) “Classification of Security Threats in Information Systems” The 5th International Conference on Ambient Systems, Networks and Technologies (ANT-2014).
2. William Stallings (2011) “Cryptography and Network Security” 6th edition. Prentice Hall
3. J.P. Anderson(1980), “Computer security threat Monitoring and surveillance” Technical Report, James P. Anderson Co, Fort Washington, PA
4. P. Barford and D. Plonka (2001), “Characteristics of network traffic flow anomalies”, Proceedings of the 1st ACM SIGCOMM Workshop on Internet Measurement,ACM New York, NY, USA, 2001, pp. 69–73.
5. A. Ghosh, A. Schwartzbard, M. Shatz (1999) Learning Program Behavior Profiles for Intrusion Detection, Proc.1st USENIX Workshop on Intrusion Detection and Network Monitoring.
6. P.E. Proctor(2001) The Practical Intrusion Detection Handbook (Prentice-Hall, Englewood Cliffs) pp. 108–111
7. S. Axelsson, Intrusion detection systems (2000)“A survey and taxonomy”, Tech. Report 99-15, Chalmers University of Technology, Department of Computer Engineering.
8. Gulshan Kumar, Krishan Kumar & Monika Sachdeva (2010) “The use of artificial intelligence based techniques for intrusion detection: a review” Published online: 4 September 2010 © Springer Science+Business Media
9. Panda, Mrutyunjaya, Ajith Abraham, and Manas Ranjan Patra (2012) “A Hybrid Intelligent Approach for Network Intrusion Detection,” Procedia Engineering 30 (2012), 1-9
10. F. Aminzadeh and M. Jamshidi (1994) “Soft Computing”. Prentice Hall.
11. Wang G, Hao J, Ma J Huang L. (2010) “A new approach to intrusion detection using Artificial Neural Networks and fuzzy clustering” Expert Systems with Applications 37 (2010) 6225–6232.
12. Christopher M. Bishop (1995) “Neural Networks for Pattern Recognition”, Oxford presses 1995.
13. Cihan H. Dagli & Pipatpong Poshyanonda (1994), “Basic artificial neural network architectures” Artificial Neural Networks for Intelligent Manufacturing (39-65) Springer Netherlands.
14. Bhavin S. and Bhushan H. T. (2012), “Artificial Neural Network based Intrusion Detection System: A Survey”, International Journal of Computer Applications 39(6):13-18.

15. J.P. Anderson(1980), "Computer security threat monitoring and surveillance" Technical Report, James P. Anderson Co, Fort Washington, PA
16. Dorothy E. Denning, D. L. Edwards, R. Jagannathan, T. F. Lunt, and P. G. Neumann (1987) A Prototype IDES—A Real-Time Intrusion Detection Expert System. Technical report, Computer Science Laboratory, SRI International.
17. Fox, K. L., Henning, R. R., Reed, J. H., and Simonian, R. (1990). A neural network approach towards intrusion detection. In Proceedings of the 13th National Computer Security Conference, 125-134.
18. A. Chittur, "Model Generation for an Intrusion Detection System Using Genetic Algorithms," High School Honors Thesis, Ossining High School, Ossining, NY, 2001.
19. W. Li, (2004) "Using Genetic Algorithm for Network Intrusion Detection". "A Genetic Algorithm Approach to Network Intrusion Detection". SANS Institute, USA.
20. Tao Xia, Guangzhi Qu, Salim Hariri & Mazin Yousif (2005), "An efficient Network Detection Method Based on Information Theory and Genetic Algorithm", Proceedings of the 24th IEEE International Performance Computing and Communications Conference, Phoenix, Arizona, USA, 2005
21. B. Abdullah, I. Abd-alghafar, Gouda I. Salama, A. Abd-alhafez, "Performance Evaluation of a Genetic Algorithm Based Approach to Network Intrusion Detection System", 2009.
22. W. Lu & I. Traore (2004), "Detecting New Forms of Network Intrusion Using Genetic Programming". Computational Intelligence, vol. 20, pp. 3, Blackwell Publishing, Malden, pp. 475-494, 2004.
23. R. H. Gong, M. Zulkernine, P. Abolmaesumi (2005), "A Software Implementation of a Genetic Algorithm Based Approach to Network Intrusion Detection", 2005
24. B.Upalhaiah, K. Anand, B. Narsimha, S. Swaraj, T. Bharat, "Genetic Algorithm Approach to Intrusion Detection System" ISSN: 0976-8491 (online) | ISSN : 2229-4333 (print), IJCST VOL3, ISSUE 1, JAN-MARCH 2012.
25. Shrinivasa K G, Saumya chandra, Sidharth Kajaria, Shilpita mukharjee, "IGIDS: Intelligent intrusion detection system using Genetic Algorithm", 978-1-4673-0126-8/11/2011 IEEE.
26. Anup Goyal, Chetan Kumar, "GA-NIDS : A genetic algorithm based network intrusion detection system",
27. Atul Kamble, "Incremental Clustering in data mining using genetic algorithm", IJCTE, Vol 2, No. 3, June, 2010
28. Shaik Akbar, Dr. J. A. chandulal, Dr. K. Nageswara Rao, G. Sudheer Kumar, "troubleshooting technique for intrusion detection sytem using genetic algorithm", IJWBC, vol 1(3), december 2011
29. Melanie Mitchell (1998), "An Introduction to Genetic Algorithms "
30. Suhail Owais, Vaclav Snasel, Pavel Kromer, Ajith abrahim,"Survey: Using genetic algorithm approach in intrusion detection system techniques", 7th computer information system and industrial management applications,2008 IEEE
31. Scott M(2004) "An introduction to genetic algorithms" Journal of Computing Sciences in Colleges Volume 20 Issue 1, October 2004 Pages 115-123.
32. V. Moraveji Hashmei, Z. Muda and W. Yassin, "Improving Intrusion Detection using Genetic Algorithm", International Technology journal 12(11) pp. 2167-2173, 2013.