International Journal of Innovation and Applied Studies
ISSN 2028-9324 Vol. 5 No. 3 Mar. 2014, pp. 233-240

http://www.issr-journals.org/ijias/

Genetic Algorithms in Intrusion Detection Systems: A Survey

Parry Gowher Majeed and Santosh Kumar

Department of Computer Science and Engineering,
Graphic Era University,
Dehradun, Uttarakhand, India

ABSTRACT: Securing the digital assets is a major concern in the present digital information era. Various tools and techniques have been researched and implemented to secure the digital assets at both individual and organizational levels. Intrusion detection systems are considered as the cornerstone of modern information security. These systems enable us to be safe from the malicious users, who intend to misuse our digital data and resources. There are different approaches, methods, and techniques employed within the field of intrusion detection. Intrusion detection based on evolutionary methods is currently a hot topic of research. Various evolutionary techniques have been successfully implemented for intrusion detection. In this paper, a survey on applications of genetic algorithms in intrusion detection systems is carried out. The paper provides an introduction to the basic concepts of intrusion detection and genetic algorithms. The generic implementation of genetic algorithms using pseudo code is presented. Pseudo code for genetic algorithm based intrusion detection method is also included for clear understanding. The paper also provides an overview of the advantages and disadvantages of genetic algorithms in general, and as applied to intrusion detection in particular. This survey will provide helpful insight into the related literature and implementation of genetic algorithms in intrusion detection systems. It will also be a good source of information for people interested in the genetic algorithms based intrusion detection systems.

KEYWORDS: Misuse Detection, Anomaly Detection, IDS Architecture, Optimization, Classification, Model Generation.

1 INTRODUCTION

In the past few decades, the computer technology has evolved at a very fast pace. This fast growth of computer technology has resulted in the transfer of more and more services to computer based systems. The dependency of more and more services on computer technology has resulted in the increase of computer related threats. With each passing day, the avoidance and detection of threats to computer technology is becoming more and more difficult. The increase in the number and severity of threats has given birth to a new field of study. Information security is the field of study dealing with security of computer systems in general. Most of the security mechanisms designed so far, try to prevent unauthorized access to system resources and data. However, it appears that such systems are not able to completely prevent intrusions into computer systems. The need is to detect intrusions efficiently, so that their impact can be realized and damages can be repaired. Also, efficient detection of the intrusions will enable security professionals to devise measures that can be used to prevent them from happening in the future. Intrusion detection systems are the tools used for prevention and detection of threats or breaches to computer systems. A lot of research has been carried out in developing and implementing new techniques ranging from basic statistical methods to highly complex evolutionary methods for intrusion detection.

The aim of this paper is to present a survey of the contributions from researchers and industry that investigate and support the use of genetic algorithms in designing intrusion detection systems.

The remainder of this paper is divided into several sections. It will start with a formal introduction to intrusion detection systems and genetic algorithms in section 2 to make readers familiar with the basic concepts needed. In the section 3, a

Corresponding Author: Parry Gowher Majeed (parry.gowher.majeed@gmail.com)

survey of the literature supporting the use of genetic algorithms for intrusion detection is presented. Section 4 details the genetic algorithms in intrusion detection. Finally, section 5 presents the conclusion derived from the survey.

2 BACKGROUND

2.1 INTRUSION DETECTION SYSTEMS

Intrusion detection systems are considered as the first line of defense in securing computer systems. They are designed to monitor and defend computer systems against intrusions. Intrusion detection systems dynamically monitor and analyze the events occurring in a system, and decide the degree of their legitimateness [1]. Intrusion detection systems are classified as network intrusion detection systems (NIDS), host intrusion detection systems (HIDS), or distributed intrusion detection systems (DIDS), based on whether an intrusion detection system monitors a network, or a host, or both [2].

Intrusion detection systems are also classified into two types on the basis of detection approach used, namely (i) misuse detection based and (ii) anomaly detection based. In misuse based intrusion detection systems, the intrusions are identified by matching collected data with a pre-specified set of signatures or by applying a set of defined rules [3]. Therefore, known intrusions are identified easily, but the problem arises with such systems when no signature exists for an intrusion. This approach has advantage of producing very low false positives. To overcome the problem of unknown intrusions, another approach to implement is anomaly detection. This approach was proposed by Denning [4] in 1987.

Anomaly based intrusion detection systems detect intrusions by analyzing deviation from expected behavior in the captured data. If the deviation crosses a certain threshold, the data is said to be anomalous. The anomaly detection approach has the capability of detecting unknown intrusions, but the major difficulty with anomaly based approach is defining what constitutes normal behavior and abnormal behavior. Another problem with the anomaly detection approach is high false positive rate [5].

Figure 1, adopted with modifications from [6], gives a generic architecture of an intrusion detection system.

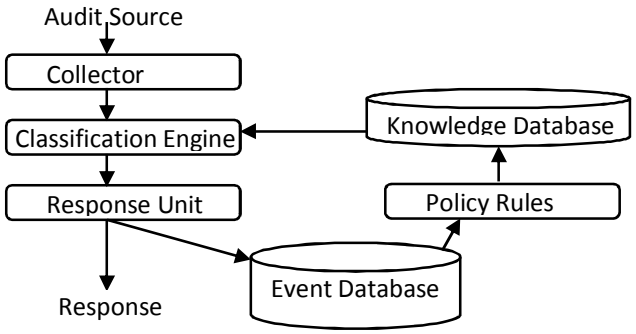

Fig. 1. Generic Architecture of Intrusion Detection System

In figure 1, the audit source represents the input to the intrusion detection system. The format of input data can be of different types depending upon the type and location of the intrusion detection system. The collector samples and pre-processes the audit source data. The data is transformed into a standard format known to the internal components of the intrusion detection system. The knowledge database contains information about attacks. The classification engine determines the legitimateness of the received data by comparing it with the attack information stored in the knowledge database. The policy rules are used to configure the response and detection of intrusion system. The response unit produces different types of responses depending upon the incoming events and their severity. The event database stores the detailed information about the events, which is used for various purposes like attack report generation, and framing new rules.

2.2 GENETIC ALGORITHMS

Li [7] describes genetic algorithm as a family of computational models based on evolution and natural selection. Bobor [8] has defined a genetic algorithm as a programming technique, which mimics biological evolution as a problem solving approach. An early work by Holland highlights the benefits of applying nature inspired adaptability function into artificial systems [9]. The genetic algorithms use techniques inspired by biological concepts like inheritance, mutation, selection, and

crossovers. The genetic algorithms are said to follow the famous "Darwinian Principle of Evolution" in functioning, which advocates the survival of the fittest among a population. Therefore, a solution obtained by applying genetic algorithms to any problem, consists of only those optimal candidate solutions which are said to satisfy a predefined fitness value [10], [11].

2.2.1 STRUCTURE OF GENETIC ALGORITHMS

Genetic algorithms are implemented as chromosome-like data structures. Figure 2 adopted from [12] depict the structure and processing in a genetic algorithm.

A genetic algorithm has many parameters, operators and processes which decide its arrival to an optimal solution. A short description of the parameters, operators and processes as depicted in figure 2, is as:

Fitness Function: The fitness function is the measure of the quality of a particular solution. The fitness function is used to determine the most optimal solution from a number of solutions in a population.

Selection: The selection process in genetic algorithms is used to select the most optimal solution determined by using the fitness function. The solutions which are not optimal are discarded.

Crossover: The crossover process in genetic algorithms is used to exchange characteristics between two different solutions. The pairs of solutions to exchange characteristics are selected randomly and keep exchanging characteristics, until a completely new generation of solutions is obtained.

Mutation: The mutation process in genetic algorithms changes some random bits in a solution. The change in the bits results in the genetic diversity of the mutated algorithms.

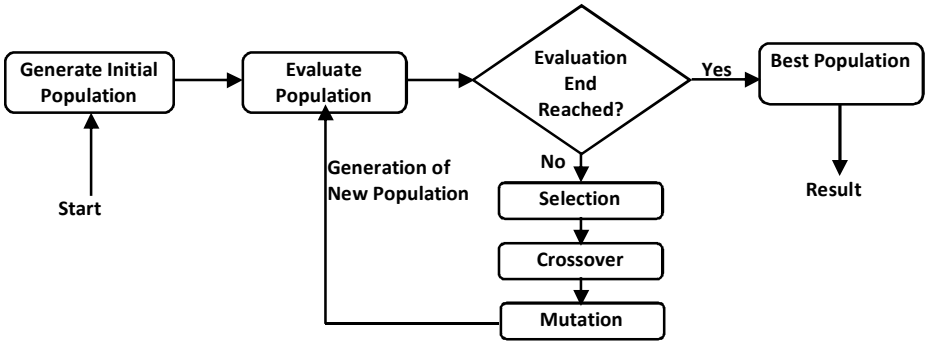

Fig. 2. Structure & Processing in a Genetic Algorithm.

2.2.2 FUNCTIONING OF GENETIC ALGORITHMS

The genetic algorithms start processing by initially selecting a random population of chromosomes. Each chromosome is composed of a finite number of genes, which is predefined in every implementation [13]. These chromosomes are the data representing the problem. This initial population is refined to a high quality population of chromosomes, where each chromosome satisfies a predefined fitness function. According to the requirements of the solution needed, different gene positions in a chromosome are encoded as numbers, bits, or characters. Each population is refined by applying mutation, crossover, inversion, and selection processes. The generic pseudo code for a genetic algorithm taken from [14] is given below for better understanding of the process:

```
InitPopulation (P)
Fitness(P)
While MaxGenerationNotReached do
        for i = 0 to xfactor do
                p1 = Selection(P)
                p2 = Selection(P)
                (o1, o2) = crossover(p1, p2)
```

```
                Crowding(p1, p2, o1, o2)
        end for
        for i = 0 to dfactor do
                p = Selection(P)
                Dropping(P)
        end for
        for i = 0 to mfactor do
                p = Selection(P)
                Mutation(p)
        end for
Fitness(P)
end while
SelectionBestIndual(P)
```

2.2.3 ADVANTAGES OF GENETIC ALGORITHMS

The various advantages of genetic algorithms are:

- Genetic algorithms possess tremendous capabilities for parallel processing.
- Genetic algorithms provide a wider solution space.
- Genetic algorithms possess easily discoverable global minima.
- Genetic algorithms are easy to modify.
- Genetic algorithms handle functions with noise efficiently.
- Genetic algorithms show high performance even in the case of multi – modal problems.
- Genetic algorithms do not need prior knowledge of the problem space.
- Genetic algorithms are least affected by the discontinuities in the problem space.
- Genetic algorithms are reliable enough not to become trapped in local minima.

2.2.4 LIMITATIONS OF GENETIC ALGORITHMS

Genetic algorithms are efficient, but in practice they have certain limitations:

- It is not always easy to find a fitness function.
- Representing a problem space in genetic algorithms is very complex.
- In many cases genetic algorithms converge prematurely to a solution.
- It is a tough task to choose the optimal parameters for a genetic algorithm.
- Genetic algorithms need to be coupled with a local searching technique for effective functioning.
- Genetic algorithms need a large number of fitness function evaluations.
- It is not easy to configure a genetic algorithm based system.

2.2.5 DIFFERENCE BETWEEN GENETIC ALGORITHMS AND CONVENTIONAL METHODS

Genetic Algorithms differ from conventional methods used for optimization. The main differences are:

- Conventional optimization methods operate on the problem parameters directly, while as genetic algorithms operate on the coded version of the problem parameters.
- Most of conventional methods operate on a single solution for producing an optimal solution, while as genetic algorithms operate on a population of solutions, selecting more optimized solutions in each iteration.
- Conventional methods usually use derivatives for evaluating the solution produced, while as genetic algorithms use a fitness function for evaluating the optimal solution produced.

- Conventional methods use deterministic transition operators, while as genetic algorithms use probabilistic transition operators.

3 LITERATURE REVIEW

Li [7] has applied genetic algorithms on both temporal and spatial network connection data to identify anomalous network behaviors. The early work regarding the application of genetic algorithms for intrusion detection is by Forrest, et. al. [15]. They have used an algorithm based on rough sets and improved genetic algorithms to improve feature selection. Crosbie and Spafford [16] have applied genetic algorithms and multi-agent technique for network anomaly detection. A combination of genetic algorithms and fuzzy data mining techniques for network intrusion detection has been proposed by Bridges and Vaughn [17]. Chittur [18] presented a genetic algorithm based model for intrusion detection, which achieved a significantly low false alarm rate. Castro and Zubin in [19] proposed a hybrid algorithm with correlation based feature selection (CFS), and employed the SVM and genetic algorithm to achieve the optimization of intrusion detection. Gome [20] used log file trace events in an off-line mode to improve the classification rules of genetic algorithm. The implementation of genetic algorithms on top of information theory to enhance intrusion detection has been proposed by Xiao, et. al. [21]. Genetic algorithms have been used for classification of Smurf attack labels in training data set, achieving a false positive rate as low as 0.2% by Goyal and Kumar [22]. Abdullah, et. al. [23] have used genetic algorithms for obtaining classification rules for intrusion detection. Ojugo, et. al. [24], have used genetic algorithms to develop rule-based intrusion detection. The fitness function has been used to evaluate the rules.

4 GENETIC ALGORITHMS IN INTRUSION DETECTION SYSTEMS

This section begins with an introduction to the working of genetic algorithms when applied to intrusion detection and an overview of an intrusion detection algorithm implemented using genetic algorithm technique. Then, the role played by genetic algorithms in intrusion detection is discussed. At the end, the advantages of implementing intrusion detection systems using genetic algorithms are presented.

The working of a genetic algorithm when applied to intrusion detection can be viewed as a sequence of following steps:

i) The packet capturing module or sniffer present in the intrusion detection system collects the information about the network traffic or logs.
ii) The intrusion detection system applies genetic algorithms to the captured data. The genetic algorithm at this stage has classification rules learned from the information collected.
iii) The intrusion detection system then applies the set of rules produced in the previous phase to the incoming traffic. Application of rules to captured data results in the population initialization, which in turn results in the creation of a new population with good qualities. This population is then evaluated and a new generation with better qualities is created. Then genetic operators are applied to the newly created generation until the most suitable individual is found.

Figure 3, provides an example of genetic algorithm implementation in intrusion detection systems:

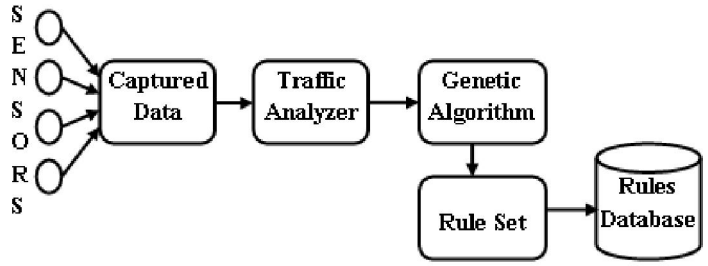

Fig. 3. Genetic Algorithm Implementation in Intrusion Detection Systems.

The working of genetic algorithms as applied to intrusion detection systems can be represented in pseudo code as:

```
initializePopulation(P)
setWeight(W)
```

```
setFitness(F)
numRules(R)
for each chromosome in P
   numAttack = 0
   numAttackNormal = 0
   for each record in T
      if the record matches the chromosome
         numAttackNormal = numAttackNormal + 1
      end if
      if the rule matches only the "condition" part
         numAttack = numAttack + 1
      end if
   end for
   calculate(F)
   if F > T
      apply the selection algorithm
      select the chromosomes into pNew
   end if
end for
for each chromosome in pNew
   apply crossover operator
   apply mutation operator
end for
if numGen not reached
   gotoline5
end if
```

Where,

P = Initial population, W = Weight value, F = Fitness function threshold, T = Training set, pNew = New population created, and numGen = Total number of generations created.

4.1 Role of Genetic Algorithms in Intrusion Detection

The subfields within the intrusion detection where genetic algorithms have been used extensively are – optimization, automatic model designing, and classification. Chittur [18] and Xiao, et. al. [21] used genetic algorithms for searching for transformation functions. Gassata [25] and Dass [26] have used a genetic algorithm for optimization purposes within the intrusion detection system. Gomez, et. al. in [20] and [27], have improved upon the application of genetic algorithm provided in [25]. Hofmann, et. al. [28] have used a genetic algorithm for optimal feature set selection and learning the structure of a radial basis function (RBF) net. Lu and Traore [29] used genetic algorithms to decide the number of clusters. Mischiattiand Neri [30], Jianet, et. al. [31], and Bankovic [32] have used genetic algorithm based classification rules as classifiers in intrusion detection systems.

A brief list of research works presenting the use of genetic algorithms in intrusion detection systems for different purposes is given in table 1:

Table 1. Fields of Application and Related Research Papers

Field of Application	List of Research Papers
Optimization	[20], [25], [26], [27], [33].
Automatic model Generation	[28], [34], [35].
Classification	[29], [30], [31], [32].

4.2 GENETIC ALGORITHM ADVANTAGES TO INTRUSION DETECTION SYSTEMS

The implementation of genetic algorithms offers many advantages to intrusion detection systems. The benefits of using genetic algorithms for intrusion detection can be summarized as:

- Genetic algorithms offer intrusion detection systems an intrinsic parallelism.
- Genetic algorithms are capable of working in multiple directions simultaneously. This makes them beneficial for analyzing the huge volumes of multi-dimensional data to be processed by an intrusion detection system.
- Genetic algorithms work with populations of solutions rather than a single solution. This makes them suitable for behavior based intrusion detection, where the behavior attributes may exhibit varying values.
- Genetic algorithms are highly re-trainable. Therefore, using genetic algorithms for intrusion detection will add to the adaptability of the system.
- Genetic algorithms evolve over time by using crossover and mutation. Property of evolving over time makes them a good choice for dynamic rule generation.

5 CONCLUSION

Intrusion detection methods based upon genetic algorithms have attracted considerable attention from the research community and the industry during the past decade. The correspondence between the requirements for building efficient intrusion detection systems and the features of genetic algorithms is the main reason behind genetic algorithms getting such an attention from the intrusion detection research community.

This survey provides an introduction to intrusion detection and genetic algorithms. The generics of genetic algorithm based intrusion detection systems are discussed. Also, the work done by different researchers in the direction of applying genetic algorithms for intrusion detection is surveyed. The paper will prove as a good starting point for newcomers to the field of genetic algorithms based intrusion detection and will also be useful for people looking for a quick review of the recent developments in the field .

REFERENCES

[1] H. Debar, M. Dacier, and A. Wespi, “Towards a Taxonomy of Intrusion Detection Systems,” *Computer Networks*, vol. . 31, no. 8, pp. 805–822, 1999.

[2] D. B. P. and M. Pels, “Host-Based Intrusion Detection Systems,” Faculty of Science, Informatics Institute, University of Amsterdam, Technical Report, 2005.

[3] Aleksandar, Kumar, and Jaideep, *Managing Cyber Threats: Issues, Approaches, and Challenges*. Springer Science + Business Media, 2005.

[4] D. E. Denning, “An Intrusion Detection Model,” *Special issue on Computer Security and Privacy*, vol. 13, no. 2, pp. 222–232, 1987.

[5] P. Garcı´a-Teodoro, J. Dı´az-Verdejo, G. Macia´-Ferna´ndez, and E. Va´zquez, “Anomaly-based Network Intrusion Detection: Techniques, Systems and Challenges,” *Computers & Security*, vol. 28, pp. 18 – 28, 2008.

[6] M. Arvidson and M. Carlbark, “Intrusion Detection Systems: Technologies, Weaknesses, and Trends,” 2003.

[7] W. Li, “Using Genetic Algorithm for Network Intrusion Detection,” in *Proceedings of the United States Department of Energy Cyber Security Group 2004 Training Conference*, 2004, pp. 1–8.

[8] V. Bobor, “Efficient Intrusion Detection System Architecture Based on Neural Networks and Genetic Algorithms.” Department of Computer and Systems Sciences, 2006.

[9] J. Holland, *Adaptation in Natural and Artificial System*. Ann Arbor, The University of Michigan Press, 1975.

[10] F. EidHebba, A. Darwish, A. E. Hassanien, and K. Tai-Hoon, “Intelligent Hybrid Anomaly Network Intrusion Detection System,” in *CCIS 265*, 2011, vol. Part I, pp. 209–218.

[11] K. G. Srinivasa, “Application of Genetic Algorithms for Detecting Anomaly in Network Intrusion Detection Systems,” *Lecture Notes of The Institute for Computer Sciences, Social Informatics & Telecommunication Engineering*, vol. 84, pp. 582–591, 2012.

[12] P. Hartmut, *Genetic and Evolutionary Algorithms: Principles, Methods and Algorithms, Genetic and Evolutionary Algorithm Toolbox*.

[13] S. Sivanandam and S. N. Deepa, *Introduction to Genetic Algorithms*. Springer-Verlag Berlin Heidelberg, 2008.

[14] V. Bapuji, R. N. Kumar, A. Goverdan, and S. Sharma, “Soft Computing and Artificial Intelligence Techniques for Intrusion Detection System,” *Networks and Complex Systems*, vol. 2, no. 4, 2012.

[15] S. Forrest, A. S. Perrelason, L. Allen, and R. Cherukur, "Self-Non-Self Discrimination in a Computer," in *Proceedings of the 1994 IEEE Symposium on Research in Security and Privacy*, 1994, pp. 202–212.
[16] M. Crosbie and E. Spafford, "Applying Genetic Algorithms to Intrusion Detection," in *Proceedings of the AAAI*, 1995.
[17] S. M. Bridges and R. B. Vaughn, "Fuzzy Data Mining and Genetic Algorithms Applied to Intrusion Detection," in *Proceedings of the National Information Systems Security Conference*, 2000, pp. 13–31.
[18] A. Chittur, "Model Generation for an Intrusion Detection System Using Genetic Algorithms," High School Honors Thesis, Ossining High School, Ossining, NY, 2001.
[19] L. De Castro and F. Von Zuben, "Learning and Optimization Using the Clonal Selection Principle," *IEEE Transactions on Evolutionary Computation*, vol. 6, no. 3, pp. 239–251, 2002.
[20] P. A. Diaz-Gome and D. F. Hougen, "Improved Off-Line Intrusion Detection Using a Genetic Algorithm," in *Proceedings of the Seventh International Conference on Enterprise Information Systems*, Miami, USA, 2005.
[21] T. Xiao, G. Qu, S. Hariri, and M. Yousif, "An Efficient Network Detection Method Based on Information Theory and Genetic Algorithm," in *Proceedings of the 24th IEEE International Performance Computing and Communications Conference*, USA, 2005.
[22] A. Goyal and C. Kumar, "GA-NIDS: A Genetic Algorithm based Network Intrusion Detection System," Electrical Engineering & Computer Science, North West University, Technical Report, 2008.
[23] B. Abdullah, Abd-algafar I., G. I. Salama, and Abd-alhafez A., "Performance Evaluation of a Genetic Algorithm Based Approach to Network Intrusion Detection System," in *Proceedings of 13th International Conference on Aerospace Sciences and Aviation Technology (ASAT-13)*, Military Technical College, Cairo, Egypt, 2009.
[24] A. A. Ojugo, A. O. Eboka, O. E. Okanta, R. E. Yora, and F. O. Aghware, "Genetic Algorithm Rule-Based Intrusion Detection System (GAIDS)," *Journal of Emerging Trends in Computing and Information Sciences*, vol. 3, no. 8, pp. 1182 – 1194, 2012.
[25] L. Mé. GASSATA, "A Genetic Algorithm as an Alternative Tool for Security Audit Trails Analysis," in *Proceedings of the 1st International Workshop on the Recent Advances in Intrusion Detection (RAID 98)*, Belgium, 1998.
[26] M. Dass, "LIDS: A Learning Intrusion Detection System," M. Sc. Dissertation, The University of Georgia, Athens, Georgia, 2003.
[27] P. A. Diaz-Gomez and D. F. Hougen, "Analysis of an Off-line Intrusion Detection System: A Case Study in Multi-objective Genetic Algorithms," in *Proceedings of the Eighteenth International Florida Artificial Intelligence Research Society Conference*, USA, 2005, pp. 822–823.
[28] A. Hofmann, C. Schmitz, and B. Sick, "Rule Extraction from Neural Networks for Intrusion detection in computer networks," in *Proceedings of IEEE International Conference on Systems, Man and Cybernetics*, 2003, vol. 2, pp. 1259–1265.
[29] W. Li, "A Genetic Algorithm Approach to Network Intrusion Detection," SANS Institute, Technical Report, 2004.
[30] M. Mischiatti and F. Neri, "Applying Local Search and Genetic Evolution in Concept Learning Systems to Detect Intrusion in Computer Networks," in *Proceedings of the 11th European Conference on Machine Learning (ECML '00)*, 2000, vol. 1810.
[31] G. Jian, L. Da-Xin, and C. Inge, "An Induction Learning Approach for Building Intrusion Detection Models using Genetic Algorithms," in *Proceedings of The 5th World Congress on Intelligent Control and Automation (WCICA 2004)*, Hangzhou, China, 2004, vol. 5, pp. 4339–4342.
[32] Z. Bankovic, D. Stepanovic, S. Bojanica, and O. Nieto-Taladriz, "Improving Network Security using Genetic Algorithm Approach," *Security of Computers & Networks*, vol. 33, pp. 438–451, 2007.
[33] P. A. Diaz-Gomez and D. F. Hougen, "A Genetic Algorithm Approach for Doing Misuse Detection in Audit Trail Files," in *Proceedings of The 15th International Conference on Computing (CIC '06)*, 2006, pp. 329–338.
[34] Q. Xu and W. Pei, "An Intrusion Detection Approach Based on Understandable Neural Network Trees," *International Journal of Computer Science and Network Security*, vol. 6, no. 11, pp. 229–234, 2006.
[35] S. Chavan, K. Shah, N. Dave, S. Mukherjee, A. Abraham, and S. Sanyal, "Adaptive Neuro-Fuzzy Intrusion Detection Systems," in *Proceedings of IEEE International Conference on Information Technology: Coding and Computing (ITCC'04)*, 2004, vol. 1, pp. 70–74.