An Intrusion Detection System Approach with Genetic Algorithm and Fuzzy Logic for Implementation of Software Networks

Mrs.Poornima B.G.* &Vinayashree**
*Assoc. Professor,
**3rd Semester, M. Tech
Computer Science and Engineering, VidyaVardaka College of Engineering,MYSURU, INDIA

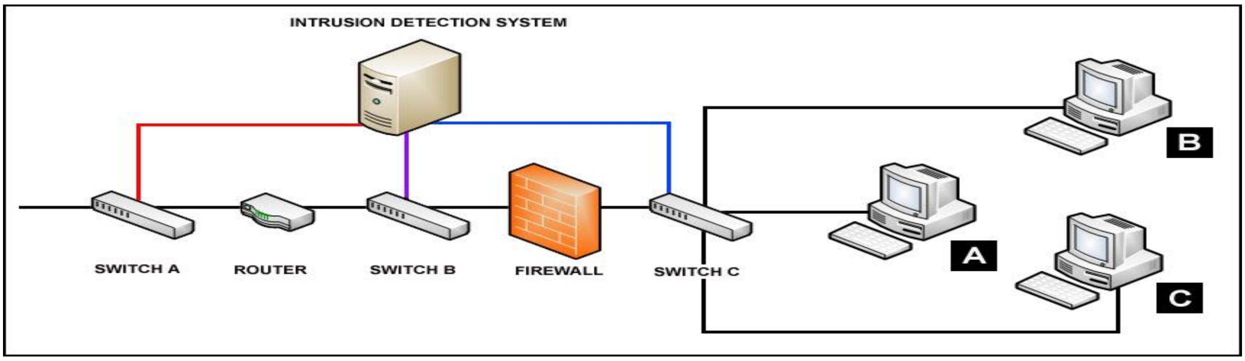
ABSTRACT:

An intrusion detection system (IDS) is a device or software application that awnings network or system activities for malicious activities or policy violations and produces reports to a management station. IDS come in many of flavours and approach the goal of detecting suspicious traffic in different ways. These days Intrusion Detection System (IDS) which is defined as a solution of system security is employed to identify the abnormal activities in a computer system or network. So far different approaches have been utilized in intrusion detections, but unluckily any of the systems is not entirely ideal. Hence, the hunt of improved method goes on. In this progression, here I am giving a survey of an Intrusion Detection System (IDS), by applying genetic algorithm (GA) and fuzzy logic to efficiently detect various types of the intrusive activities within a network. Different soft computing based approaches have been proposed to detect computer network attacks. This paper presents a survey of genetic algorithm (GA) based approach to network intrusion detection. The genetic algorithm is employed to derive a set of classification rules from network audit data, and the support-confidence framework is utilized as fitness function to judge the quality of each rule.

1. Introduction

The Internet and local area networks are expanding at an amazing rate in recent years. While we are benefiting from the convenience that the new technology has brought us, computer systems are exposed to increasing security threats that originate externally or internally. Different but complementary technologies have been developed and deployed to protect organizations' computer systems against network attacks, for example, anti-virus software, firewall, message encryption, secured network protocols, password protection, and so on. Despite different protection mechanisms, it is nearly impossible to have a completely secured system. Therefore, intrusion detection is becoming an increasingly important technology that monitors network traffic and identifies network intrusions such as anomalous network behaviours, unauthorized network access, and malicious attacks to computer systems [7].

There are two general categories of intrusion detection systems (IDSs): misuse detection and anomaly detection [8]. Misuse detection systems detect intruders with known patterns, and anomaly detection systems identify deviations from normal network behaviours and alert for potential unknown attacks. Some IDSs integrate both misuse and anomaly detection and form hybrid detection systems. The IDS scan also be classified into two categories depending on where they look for intrusions. A host-based IDS monitors activities associated with a particular host, and a network-based IDS listens to network traffic. A number of soft computing based approaches have been proposed for detecting network intrusions [1, 2, 3, and 4].



Figur1. Intrusion detection system

Soft computing refers to a group of techniques that exploit the tolerance for imprecision, uncertainty, partial truth, and approximation to achieve robustness and low solution cost. The principle constituents of soft computing are Fuzzy Logic (FL),Artificial Neural Networks (ANNs), Probabilistic Reasoning (PR), and Genetic Algorithms (GAs).Intrusion detection is designed to monitor the malicious activities ([1], [2], [3] and[4]) occurring in a computer systemor network inside or outside and analysing them for signs of possible incidents, which are violations or forthcomingthreats of violation of computer security policies, acceptable utilized policies, or standard security practices. Intrusionincidents to computer systems are increasing because of the commercialization of the internet and local networksand new automated hacking tools. Computer systems are turning out to be more and more susceptible to attack, due toits extended network connectivity.

Intrusion Detection Systems (IDS) are primarily focused on identifying possible incidents, logging informationabout them, attempting to stop them, and reporting them to security administrators in real-time, or near real-time,and those that process audit data with some delay (non-real-time). The latter approach would in turn delay the time ofdetection. In addition, organizations use IDSs for other purposes, such as identifying problems with security policies,documenting existing threats, and deterring individuals from violating security policies. IDSs have become a necessaryaddition to the security infrastructure of nearly every organization.

2. Genetic Algorithms

Genetic algorithms [3, 10] employ metaphor from biology and genetics to iteratively evolve a population of initial individuals to a population of high quality individuals, where each individual represents a solution of the problem to be solved and is composed of a fixed number of genes. The number of possible values of each gene is called the cardinality of the gene. Figure 1 illustrates the operation of a general genetic algorithm. The operation starts from an initial population of randomly generated individuals. Then the population is evolved for a number of generations and the qualities of the individuals are gradually improved. During each generation, three basic genetic operators are sequentially applied to each individual with certain probabilities, i.e., selection, crossover, and mutation. First, a number of best-fit individuals are selected based on a user-defined fitness function. The remaining individuals are discarded. Next, a number of individuals are selected and paired with each other.

Genetic Algorithm Incorporating the concept of evolutionary genetic algorithm to produce a set of rules which can be applied on network testing set can help in producing results in detection of intrusions in network. Various researchers have adopted genetic algorithm in different ways for detection of intrusions.

A genetic algorithm is a searching technique used to find out solution to problems which can be illustrated in numerous problem dependent ways. The effectiveness of the solution is evaluated by determining its fitness function. It finds solution space for an optimal solution to a problem. The basic characteristic of the genetic algorithm is how the search is performed. The algorithm produces a population of possible solutions for a problem and let them evolve over multiple generations for finding better solutions.

Genetic algorithm process commence with a set of potential solutions (chromosomes) which comprises a population, are generated or selected on random basis. These chromosomes evolve during various generations producing new off-springs by using techniques like crossover and mutation.

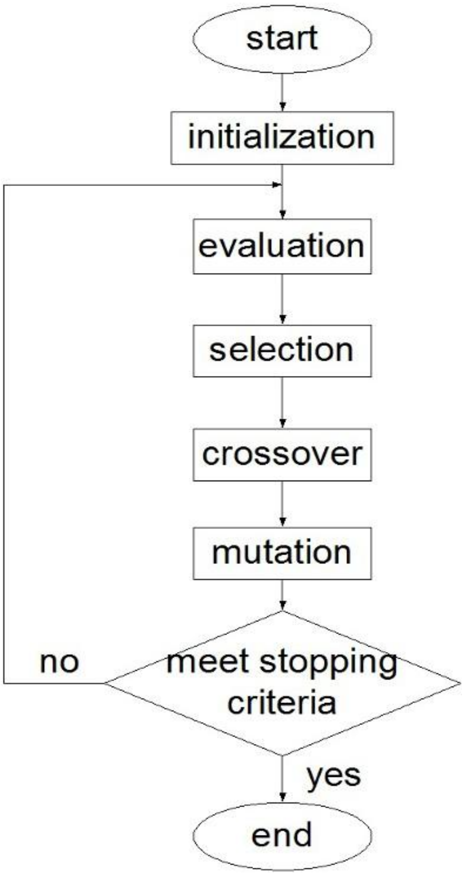


Figure 1. The operation of a generic GA.

When a GA is used for problem-solving, three factors will have impact on the effectiveness of the algorithm, they are: 1) the selection of fitness function; 2) the representation of individuals; and 3) the values of the GA parameters. The determination of these factors often depends on applications. In our implementation for network intrusion detection, the support-confidence framework was used as fitness function, a simple GA (rather than GP) was employed to represent and derive rules, and appropriate GA parameters, including selection rate, crossing over style, mutation rate, etc. were chosen based on a large number of experiments.

3. Related Work

This section briefly summarizes some of the applications of soft computing techniques for intrusion detection. However, a number of GA based IDSs are discussed in the later part of the paper in order to compare and contrast those work with our work. GAs and GP have been used for network intrusion detection in different ways. Some approaches directly use GAs to derive the classification rules [2, 5, 6, 10], while some others use different AI methods for acquisition of rules, where GAs are used to select appropriate features or to determine the optimal parameters of some functions [1, 5].

Li [5] propose a GA-based method to detect anomalous network behaviours. Both quantitative and categorical features of network data are included when deriving classification rules using GA. The inclusion of quantitative features may lead to increased detection rates. However, no experimental results are available yet.Hassan [4], Baruah ([5], [6]), Neog and Sut [9] have forwarded an extended definition of fuzzy set which enables usto define the complement of a fuzzy set. Our proposed system agrees with them as this new definition satisfies all the properties regarding the complement of a fuzzy set.

Gong [2] presented an implementation of GA based approach to Network Intrusion Detection using GA and showedsoftware implementation. The approach derived a set of classification rules and utilizes a support-confidenceframework to judge fitness function.

Xia, Hariri and Yousif [3] used GA to detect anomalous network behaviours based on information theory ([9],[11]). Some network features can be identified with network attacks based on mutual information between networkfeatures and type of intrusions and then using these features a linear structure rule and also a GA is derived. Theapproach of using mutual information and resulting linear rule seems very effective because of the reduced complexityand higher detection rate. The only problem is that it considered only the discrete features.

Abdullah [6] showed a GA based performance evaluation algorithm to network intrusion detection. The approachuses information theory for filtering the traffic data.

Lu and Traore used historical network dataset using GP to derive a set of classification. They usedsupport-confidence framework as the fitness function and accurately classified several network intrusions. But their useof genetic programming made the implementation procedure very difficult and also for training procedure more dataand time is required.

4. INTRUSION DETECTION OVERVIEW

The following sections give a short overview of networking attacks, classifications and various components ofIntrusion Detection System.

A. NETWORKING ATTACKS

This section is an overview of the four major categories of networking attacks. Every attack on a network cancomfortably be placed into one of these groups [7] –

1.Denial of Service (DoS):A DoS attack is a type of attack in which the hacker makes a computing or memoryresources too busy or too full to serve legitimate networking requests and hence denying users access to amachine e.g. apache, ping of death, back, mail bomb, UDP storm etc. are all DoS attacks.

2. Remote to User Attacks (R2L): A remote to user attack is an attack in which a user sends packets to a machineover the internet, which he/she does not have access to in order to expose the machines vulnerabilities andexploit privileges which a local user would have on the computer e.g. guest, xnsnoop,sendmaildictionary etc.

3. User to Root Attacks (U2R): These attacks are exploitations in which the hacker starts off on the system with anormal user account and attempts to abuse vulnerabilities in the system in order to gain super user privileges e.g.perl, xterm.

4.Probing: Probing is an attack in which the hacker scans a machine or a networking device in order to determineweaknesses or vulnerabilities that may later be exploited so as to compromise the system. This technique iscommonly used in data mining e.g. saint, portsweep, mscan, nmap etc.

B. CLASSIFICATION OF INTRUSION DETECTION

Intrusions Detection can be classified into two main categories. They are as follow:

1.Host Based Intrusion Detection: HIDSs evaluate information found on a single or multiple host systems,including contents of operating systems, system and application files ([11], [16]).

2.Network Based Intrusion Detection: NIDSs evaluate information captured from network communications,analysing the stream of packets which travel across the network ([11], [16]).

5. A GA-Based IDS

The proposed GA-based intrusion detection approach contains two modules where each works in a different stage. In the training stage, a set of classification rules are generated from network audit data using the GA in an offline environment. In the intrusion detection stage, the generated rules are used to classify incoming network connections in the real time environment. Once the rules are generated, theintrusion detection is simple and efficient. In thefollowing sections, we focus our discussions on deriving the set of rules using GA.

A. GENETIC ALGORITHM OVERVIEW

A Genetic Algorithm (GA) is a programming technique that uses 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 candidatesolutions towards a predefined fitness [12].

The proposed GA based intrusion detection system contains two modules where each works in a different stage. Inthe training stage, a set of classification rules are generated from network audit data using the GA in an offlineenvironment. In the intrusion detection stage, the generated rules are used to classify incoming network connections inthe real-time environment. Once the rules are generated, the intrusion detection system becomes simple, experiencedand efficient one.

B. FUZZY LOGIC

It has been shown that a fuzzy number [a, b, c] is defined with reference to a membership functionμ(x) lying between 0 and 1, a ≤ x ≤ c. Further, he has extended this definition in the following way. Let μ1(x) and μ2(x)be two functions, 0 ≤ μ2(x) ≤ μ1(x) ≤ 1. He has concluded μ1(x) the fuzzy membership function, and μ2(x) a referencefunction, such that (μ1(x) – μ2(x)) is the fuzzy membership value for any x. Finally he has characterized such a fuzzynumber by {x, μ1(x), μ2(x); x ∈Ω}.

The complement of μxis always counted from the ground level in Zadehian's theory [10], whereas it actuallycounted from the level if it is not as zero that is the surface value is not always zero. If other than zero, the problemarises and then we have to count the membership value from the surface for the complement of μx. Thus I couldconclude the following statement – Complement of μx = 1 for the entire levelMembership value for the complement of μx = 1- μx. My system forwarded a definition of complement of an extended fuzzy set where the fuzzy reference function is notalways zero. The definition of complement of a fuzzy set proposed by Hassan [4], Baruah ([7], [8]), Neog and Sut [9]could be seen a particular case of what I am giving. I shall use Baruah's definition of the complement of a normal fuzzyset in my article.

5.1. Data Representation

Several network features have higher possibilities to be involved in network intrusions [7, 9]. In our approach, seven of those features are selected from the network audit data to compose a classification rule.

5.2. Fitness Function

To determine the fitness of a rule, the support confidence framework [8] is used. If a rule is represented as if A then B, then the fitness of the rule is determined using following equations:support = |A and B| / N

confidence = |A and B| / |A|

fitness = w1 * support + w2 * confidence

Here, N is the total number of network connections in the audit data, |A| stands for the number of network connections matching the condition A, and |A and B| is the number of network connections that matches the rule if A then B. The weights w1 and w2 are used to control the balancebetween the two terms and have the default values of w1=0.2 and w2=0.8.s

Intrusion detection systems are an asset to effective security implemented on networked systems. It is basically a set of techniques that are used to detect suspicious activity at network and host level. It is generally categorized into two basic categories: Signature based and Anomaly detection systems.

Generally, an intrusion detection system captures data from network and detects anomalies in it by applying its specified rules. IDS can possess different capabilities which depend upon how complex and sophisticated the components are.

NIDS (Network-based IDS) are intrusion detection systems that acquire data packets roaming on network media and compare them to the database of signatures.

HIDS (Host-based IDS) are intrusion detection systems that are installed as agents on host and look into systems and application log files to capture any intruder activity.

6. *COMPONENTS OF INTRUSION DETECTION SYSTEM*

An intrusion detection system normally consists of three functional components [17]. The first component of anintrusion detection system, also known as the event generator, is a ***data source***. Data sources can be categorized intofour categories namely Host-based monitors, Network-based monitors, Application-based monitors and Target-basedmonitors.

The second component of an intrusion detection system is known as the ***analysis engine***. This component takesinformation from the data source and examines the data for symptoms of attacks or other policy violations. Theanalysis engine can use one or both of the following analysis approaches:

1. **Misuse/Signature-Based Detection:** This type of detection engine detects intrusions that follow Ill-knownpatterns of attacks (or signatures) that exploit known software vulnerabilities ([18], [19]). The main limitation ofthis approach is that it only looks for the known weaknesses and may not care about detecting unknown futureintrusions [20].

2. **Anomaly/Statistical Detection:** An anomaly based detection engine will search for something rare or unusual [20]. They analyses system event streams, using statistical techniques to find patterns of activity that appear tobe abnormal. The primary disadvantages of this system are that they are highly expensive and they canrecognize an intrusive behaviour as normal behaviour because of insufficient data.

The third component of an intrusion detection system is the ***response manager***. In basic terms, the responsemanager will only act when inaccuracies (possible intrusion attacks) are found on the system, by informingsomeone or something in the form of a response.

Conclusion

In this paper, a method of applying genetic algorithms for network intrusion detection is presented. A software is implemented for the presented method, and its architecture and operations are described in detail using high level class diagram and pseudo-code. One of the major advantages of this technique is due to the fact that in the real world, the types of intrusions change and become complicated very rapidly. The proposed detection system can upload and update new rules to the systems as the new intrusions become known. Therefore, it is cost effective and adaptive.

A GA is used to derive a set of classification rules from network audit data. Seven network features including both categorical and quantitative data fields were used when encoding and deriving the rules. The support-confidence framework, is used to select the appropriate rules. Depending on the selection of fitness function weight values, the generated rules can be used to either generally detect network intrusions or precisely classify the types of intrusions.

REFERENCES

[1] S. M. Bridges and R. B. Vaughn, “Fuzzy Data Mining And Genetic Algorithms Applied To Intrusion Detection”, *Proceedings of 12th Annual CanadianInformation Technology Security Symposium*, pp. 109- 122, 2000.

[2] A. Chittur, “Model Generation for an Intrusion Detection System Using Genetic Algorithms”, http:// www1.cs.columbia.edu/ids/publications/gaids-thesis 01.pdf (accessed in January 2005).

[3] M. Crosbie and E. Spafford, “Applying Genetic Programming to Intrusion Detection”, *Proceedings ofthe AAAI Fall Symposium*, 1995

[4] D. Dasgupta and F. A. Gonzalez, “An Intelligent Decision Support System for Intrusion Detection and Response”, *MMM-ACNS, Lecture Notes in Computer Science*, vol. *2052*, pp. 1-14, 2001.

[5] W. Li, “A Genetic Algorithm Approach to Network Intrusion Detection”, SANS Institute, USA, 2004

[6] W. Lu and I. Traore, “Detecting New Forms of Network Intrusion Using Genetic Programming”, *Computational Intelligence,* vol. 20, pp. 3, Blackwell Publishing, Malden, pp. 475-494, 2004.

[7] A. Adetoye, A. Choi, M. Md. Arshad, and O. Soretire, “Network Intrusion Detection & Response System”, Group Report, September 2003, http://www.cs.ucl.ac.uk/teaching/dcnds/group-reports /2003/2003-hailes-b.pdf (accessed in January 2005).

[8] B. Mukherjee, L. T. Heberlein, and K. N. Levitt, “Network intrusion detection”, *IEEE Network*, 8(3): 26-41, May/June 1994.

[9] T. Xiao, G. Qu, S. Hariri, and M. Yousif, “An Efficient Network Intrusion Detection Method Based on Information Theory and Genetic Algorithm”, *Proceedings of the 24th IEEE InternationalPerformance Computing and Communications Conference (IPCCC ‘05),* Phoenix, AZ, USA. 2005.

[10] H. Pohlheim, “Genetic and Evolutionary Algorithms: Principles, Methods and Algorithms”,

http://www.geatbx.com/docu/index.html (accessed in January 2005).

[11] Hemanta K. Baruah, “Towards Forming A Field Of Fuzzy Sets”, International Journal of Energy, Information and Communications, Vol. 2, Issue 1, pp. 16-20, 2011.

[12] Hemanta K. Baruah, “The Theory of Fuzzy Sets: Beliefs and Realities”, International Journal of Energy, Information and Communications, Vol. 2, Issue 2, pp. 1-22, 2011.

[13] TridivJyotiNeog, Dushmanta Kumar Sut, “Complement of an Extended Fuzzy Set”, International Journal of Computer Applications, Vol. 29, No.3, pp. 39-45, 2011.

[14] Zadeh L A, “Fuzzy Sets”, Information and Control, Vol.8, pp. 338-353, 1965.

[15] S. Kumar, E. Spafford, “A Software architecture to Support Misuse Intrusion Detection”, in the 18th National Information Security Conference, pp. 194-204, 1995.

[16] K. Ilgun, R. Kemmerer, P. A. Porras, “State Transition Analysis: A Rule-Based Intrusion Detection Approach”, IEEE Transaction on Software Engineering, pp. 181-199, 1995.

[27] S. Kumar, “Classification and Detection of Computer Intrusions”, Purdue University, 1995.

[18]Ren Hui Gong, Mohammad Zulkernine, PurangAbolmaesumi, “A Software Implementation of a Genetic Algorithm Based Approach to Network Intrusion Detection” 0-7695-2294-7/05 $20.00 © 2005 IEEE

[19]Mostaque Md. Morshedur Hassan, “Network Intrusion Detection System Using Genetic Algorithm and Fuzzy Logic” ISSN(Online): 2320-9801 ISSN (Print): 2320-9798 IJIRCCE