ISSN (Online) : 2278-1021
ISSN (Print) : 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 3, Issue 5, May 2014

Recent trend in Intrusion detection using Fuzzy-Genetic algorithm

Swati Sharma[1], Santosh Kumar[2], Mandeep Kaur[3]

Graphic Era University, Dehradun, India[1,2,3]

Abstract: Computer networks have expanded significantly in use and this makes them more vulnerable to attacks. It is really important to secure the data from any intrusive attacks so intrusion detection is really very helpful in the field of computer network security. Intrusion detection is the act of detecting unwanted traffic on a network. Many current intrusion detection systems are unable to find unknown attacks. A no. of GA and fuzzy logic based approaches are used for detecting network intrusions. This paper presents a survey of these approaches in intrusion detection with advantages. KDD cup data used in every technique which have information of computer networks during normal and intrusive behavior. It contains basically four categories of attacks. GA is used to optimization purpose and fuzzy logic work on approximation rather than precise values. NSLKDD is an advance version of KDD cup data set.

Keywords: Anomaly detection, Fitness function, Nsl Kdd, Network attacks

I. INTRODUCTION

Information system security is important in this computer age. It goes without saying that information management is crucial for the survival of any firm. Enterprises, for example, depend on information to run their businesses, which is increasing constantly. Hence, there is a need to ensure its security, in terms of confidentiality, integrity, and availability, to maintain a competitive edge over other businesses. It is important to note that most of the attacks are form insider who misuses their privilege. So it is therefore Intrusion detection is the act of detecting unwanted traffic on a network. IDS can be software or a physical appliance that monitors network traffic in order to detect unwanted activity and events such as illegal and malicious traffic, traffic that breaks security policy, and traffic that breaks acceptable use policies. Several types of IDS technologies exist due to the variance of network configurations. Each type has many advantages and disadvantage in detection, configuration, and cost. An IDS that looks at network traffic and detects data that is generally abnormal like incorrect, not valid is called anomaly-based detection. This method is helpful for detecting unwanted traffic that is not specifically known. For instance, an anomaly-based detect that an Internet protocol (IP) packet is malformed. IDS uses signature-based detection based on known traffic data to analyze unwanted traffic. This type of detection is easy to configure and fast. However, an attacker can modify an attack to bypass it undetectable by a signature based IDS. Still, signature based detection limited in its detection capability.

II. PROBLEM IN CURRENT IDS

Current ids is suffering from so many problems, there are lots of ids tools available in the market and the internet also but the problem is same they are not fully capable to detect every kind of attacks. Snort is most popular intrusion detection tool. Snort intrusion detection and prevention system is an open source software tool. It can be used for real time analysis of traffic over a computer network. In addition to packet payload analysis, it can analyze protocols, and can detect different type of attacks.

- First, intrusion detection depends on monitored data (audit data) in some way. When an IDS is used in the anomaly detection mode, it learns from the monitored data to detect attacks. But this monitored data is not 100% sure about the attacks, it only shows the probability of attack [2].
- Second, false positive alarm need to be decreased.
- Third, IDS uses additional resource in the system even when there is no intrusion detecting because IDS has to be run all the time.
- Fourth, speed of IDS[4].
- Fifths, accuracy.

III. RELATED WORK

In 2013 Mostaque Md. Morshedur Hassan [3] has presented methodologies and good fuzzy classifier using GA, he also presented challenges in IDS. Author proposed the new definition of fuzzy set where he described the fuzzy membership value and fuzzy membership function for the complement of a fuzzy set are two different concepts because the surface value is not always counted from ground level. The advantage of this technique helps in reducing the false alarm rate in IDS.

In 2012 N. Wattanapongsakorn et al [19] presented a network based intrusion detection and prevention system. He implemented IDPS using different machine learning algorithms like decision trees, ripple rule, random forest, and Bayesian network and tested on an online network environment. For protecting network system he used iptables to block or drop the packet. The advantage of this approach is that it allows users to select more than one algorithm to improve the performance of intrusion detection.

In 2015 Amin Einipour [5] has presented the combination of fuzzy system and particle swarm optimization (PSO) algorithm. Fuzzy rules are used because of their

ISSN (Online) : 2278-1021
ISSN (Print) : 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 3, Issue 5, May 2014

interpretability by human experts and PSO is employed as meta-heuristic algorithm to optimize the fuzzy rules. The advantage of this technique is high reliability and adequate interpretability which are features of data mining techniques.

In 2012 K. S. Anil Kumar et al [11] has presented an adaptive GA model for ID which consists of K-Mean clustering algorithm for classifying the data by analyzing it on the basis of their attributes. He used GA and neural network for making rules, which learn, differentiate and extract underlying correlations and sort out the pattern of data fed. The advantage of this model is that it helps in achieving robust architecture. K-Means, GA, and neural networks have some deficiencies as individual techniques, but they can be rectified by mixing them appropriately. In 2013 P. Jongsuebsuk et al [18] proposes a fuzzy genetic algorithm for network intrusion detection system. He considered both KDD99 dataset and online data set, online dataset has a more recent behavior of network activities than those in the KDD99. The advantage of this method is it detects network attacks in real time within 2-3 seconds after the data arrive at detection system.

In 2012 B. Uppalaish et al [12] implemented GA and trained it on KDDCUP99 dataset to generate rules for IDS to identify and classify different types of attacks connections. The advantage of this approach is that it provides high rate of rule set for detecting different types of attacks.

In 2012 Shiyong Li et al [13] proposed an intrusion detection method with the use of GA and rough set. Rough set is a kind of new mathematical tool to describe incomplete and uncertain information. GA removes the deficiency rough set. The advantage of this approach is that the system uses minimum rules to guide the IDS fast and accurately in the case of the vast amount of information.

In 2012 Shaokun Liu et al [15] presented an improved GA to solve the optimization problem and also introduced the greedy algorithm for better improvement. The advantage of this method is the performance of improved GA better than simple GA. In 2011 Madhuri Agravat et al [16] described two objectives fuzzy genetic based learning algorithms. Two objectives are:-

- Maximize detection rate
- Contain minimum no. of rules with low false.
- These two objectives combined into single scalar fitness function and then the genetic algorithm applied to them. The advantage of this scheme is that it will reduce the search space of finding rules for new patterns and also takes lesser CPU time.

In 2009 Wang Yunwu [20] analysed the current situation of IDS by using genetic based fuzzy system algorithm. He presented a genetic fuzzy expert system (GFES) in which he applied association rules together with fuzzy logic to classify the data. The advantage of this is it needs less fuzzy rules to achieve a certain high rate of recognition and classification. Fuzzy rules do not need any luminous knowledge for generating rules.

TABLE 1 FOLLOWING ARE VARIOUS TECHNIQUES WITH PROPER CITATION , YEAR AND ADVANTAGES

Citation	Year	Technique	Advantage
[11]	2014	Adaptive GA	Robust architecture
[3]	2013	Fuzzy classifier using GA	Reduce false alarm
[14]	2013	GA	Comparison of different techniques given
[18]	2013	Fuzzy genetic algorithm	It detects in real time within 2-3sec
[5]	2012	Fuzzy system and particle swarm optimization (PSO)	High reliability
[12]	2012	GA	High rate of rule
[13]	2012	GA and rough set	System uses minimum rules
[15]	2012	Improved GA in multi agent	Better performance than simple GA
[19]	2012	Intrusion detection and prevention techniques	Allow user to select more than one algorithm for ID
[16]	2011	Fuzzy genetic algorithm	Reduce search space of finding rules
[20]	2009	Genetic based fuzzy system	Less fuzzy rules to achieve high rate of recognition and classification
[23]	2009	Artificial neural network	Help in detecting probe attacks
[21]	2008	Fuzzy genetic approach	Superior performance than other GA based approach
[24]	2007	collaborative intrusion detection system (CIDS)	secure against basic probe-response attacks

IV. THE DATASET USED

Every researcher use offline dataset to implement their algorithm such as DARPA1998 data or KDD CUP 99 data. In 1998 MIT Lincoln Labs managed DARPA Intrusion Detection, Evaluation Program [6]. A standard set of audit data which include a wide range of network intrusions simulated in a military network area was provided. The aim was to survey and assess research in intrusion detection. A connection is a sequence of TCP packets, between which data flows from a source IP address to a target IP address under some well-defined protocol [6]. Each connection mark as a normal or attack with specified attack type. There are some issues in KDDCUP99 data set, to solve these issues NSL KDD new data set produced. It solves the many inherent problems of KDDCUP99 like [22]:-

- It does not include redundant record in data.
- The number of selected records from each difficulty level group is inversely proportional to the percentage of records in the original KDD data set.
- The numbers of record in train and test data set are reasonable.
- There are some issues in KDDCUP99 data set, to solve these issues NSL KDD new data set produced. It solves the many inherent problems of KDDCUP99 like [22]:-
- It does not include redundant record in data.
- The number of selected records from each difficulty level group is inversely proportional to the percentage of records in the original KDD data set.
- The numbers of record in train and test data set are reasonable.

V NETWORK ATTACKS

There are basically four network attack found in network and any one of other attacks can be categories in these four attacks [2].

- Denial of Service (DoS): It is an attack which blocks the availability of computer systems or services. It comes in a variety of forms, for example, consumption of scarce, limited, or non-renewable resources, destruction or alteration of configuration

ISSN (Online) : 2278-1021
ISSN (Print) : 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 3, Issue 5, May 2014

information, physical destruction or alteration of network components [7].

- Probing: It is an attack in which attacker scans and determines the weaknesses or the vulnerabilities in machine or network device that could be later useful for attacker, for example port scanning, mscan, nmap etc.
- User to Root Attacks (U2R): It is an attack in which hacker starts with a normal account on a system and then try to exploit the super user control for misuse purpose, for example Perl, xterm.
- Remote to User Attacks (R2L): It is a remote to user attack in which remote user sends packets to system over the network, which remote user does not has right in order to access the system and examine the privilege access which local user has on system, for example xlock, guest, phf, xnsnoop ,sendmail dictionary etc.

V. OVERVIEW OF GENETIC ALGORITHM

Genetic algorithms (GA) are search algorithms based on the principles of natural selection and genetics. The aim of development of GAs is developing a system as robust and as adaptable to the environment as the natural systems [4]. Genetic algorithms are search procedures often used for optimization problems. In this algorithm an initial population of chromosomes is generated randomly where each chromosome represents a possible solution to the problem (a set of parameters) [2]. From each chromosome different positions are encoded as, characters, bits or no.s. These positions could be known as genes. Goodness of each chromosome calculated by evaluation function, according to the desired solution; this function is known as "Fitness Function" [3]. It holds three phases after calculating fitness function i.e. selection, crossover, and mutation. In selection it selects most optimal solution of a problem calculated by using fitness function. The selected chromosomes are called parents. After selection phase crossover phase comes in which characteristics of different parent chromosomes exchange and they produce offspring, there are various methods for crossover, for example N-point crossover, uniform crossover etc. Mutation involves flipping of one or more bits of chromosomes and then evaluated using some fitness criteria. After termination chromosomes having the highest fitness function called the best solution of the problem. Mutation maintains diversity in the population [10]. Genetic algorithm from other algorithm because it implemented at machine code level, it is fast to detect in real time [8]. Some of its good properties, e.g. robust to noise, no gradient information is needed to find global optimal or sub-optimal solution, self-learning capabilities made it best approach [4]

VI. INTRUSION DETECTION USING GENETIC ALGORITHM

The proposed GA based intrusion detection system holds two modules where each acts in a different stage. Using GA a set of classification rules is produced in the training stage, from network audit data using in an offline background. The generated rules are employed in intrusion detection phase to classify incoming network connections in the real-time environment. When the rules are generated, the intrusion detection system becomes simple, experienced and efficient one [3].

Effectiveness of genetic algorithm depends on three factors:-

- Genetic algorithm parameter:-
  - Population size
  - Mutation methods
  - Recombination methods
  - Parent selection methods

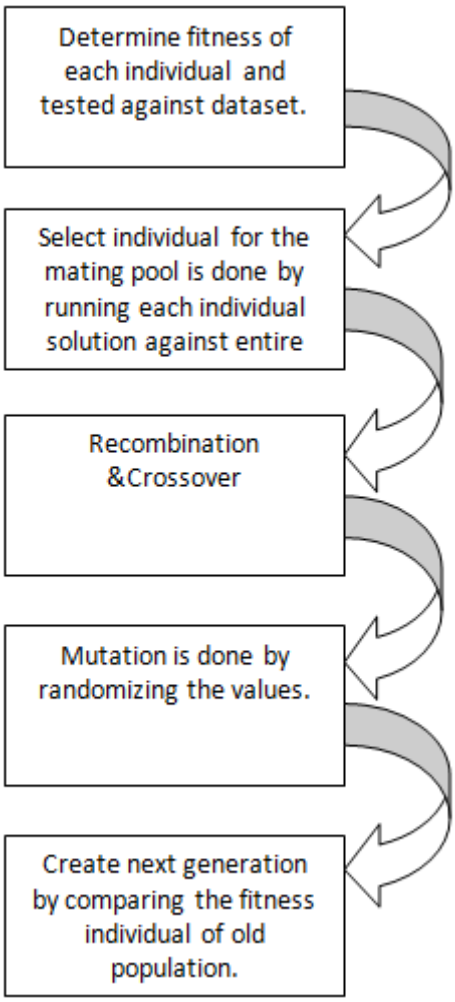


Fig. 1 Implementation phases of GA [25]

Fitness function:-

Genetic algorithm to identify the type of attack connects, the algorithm used different features in network connection to generate classification rule set, they used the fitness function given by the formula:-

$F= a/A – b/B$

Where an A= total of attacks, B= normal connections in the population, a = no. of attack connections the individual correctly classified, b = no. of normal connection a network correctly classified. GA to detect DOS and probe type of attack they used a fitness function.

$Fitness = f(x)/f(sum)$

Where f(x) is the fitness of an entity and f(sum) is the total fitness of all entities genetic algorithm cannot be done

ISSN (Online) : 2278-1021
ISSN (Print) : 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 3, Issue 5, May 2014

without selection process which depends mainly on fitness value that obtained using fitness function.

Representation of individuals:-

- Canonical GA- $\vec{a} \in \{0,1\}^l$
- Genotype space- $\{0,1\}^l$

Problem specific representation

VII. ADVANTAGE OF GENETIC ALGORITHM

There are many advantages of GA few of them are below[14, 21]:-

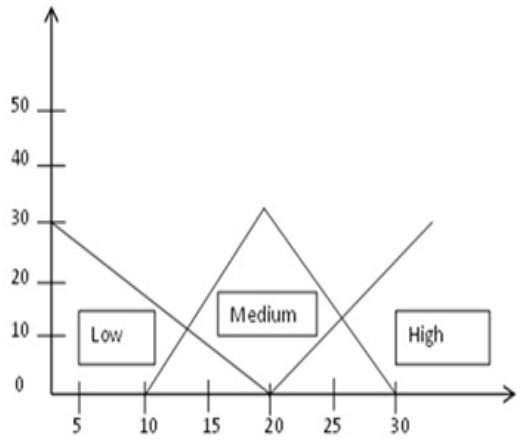
- Easily GA are parallel, they have multiple offspring if any one of them fails in providing required solution they can eliminate them and continues their work with other offspring's providing them a greater chance in finding optimal solutions.
- Adaptable system which can evolve new rules for IDS. This is useful in detecting new attacks.
- Work on populations candidate solutions rather than the single candidate solution, this allows GA to cope well with attribute interactions and avoid stuck in local maxima.
- Parallelism allows evaluate multiple schemas at once and GA is suited for solving problems whose space of potential solutions is huge to search exhaustively reasonable amount of time.

IX . OVERVIEW OF FUZZY CONCEPT

Fuzzy Logic introduced by Zadeh (1965) gives us a language, with syntax and local semantics, in which we can translate our qualitative knowledge [8]. Fuzzy logic is derived from fuzzy set theory where evaluation is approximation rather than precise values. In real world quantities which seems accurate, crisp and deterministic, but actually they are not so. They possess uncertainty which arises from imprecision. This can be represented in the form of membership functions, there are various ways to assign membership value to a fuzzy variable, for example, intuition, inference, genetic algorithm etc. fuzzy system made up of three elements [1]:-

- Fuzzifier: It converts the crisp value to linguistic value by calculating membership function.
- Inference engine: Fuzzy rule based system, fuzzy model and fuzzy expert systems are known as inference engine. The main work of this is decision making, it uses "IF THEN" rules that convert the any input to fuzzy output.
- Defuzzifier: It is opposite of fuzzifier, it converts the fuzzy input to crisp output with the help of membership functions

Membership function defines fuzziness in fuzzy set. Fuzzy main characteristic is the robustness of its interpolative reasoning mechanism [8]. Fuzzy allow partial membership means a membership value can be a part of one or more fuzzy set. It ranges between interval [0, 1] membership value 0 means it is not a part of fuzzy set while value 1 means it is completely belong to entire set [1]. As shown in fig y-axis is the degree of membership in the set and the x-axis is the value of quantitative variables. The possible values of the fuzzy variable are the fuzzy sets low, medium and high. In fuzzy logic the degree of membership can be between 0 & 1. The value 15 is a member of the set low to the degree 7 and member of the medium to the degree 18.



X. CONCLUSION

In this paper, we have provided a survey of intrusion detection using fuzzy logic and GA. A brief overview of intrusion detection system, genetic algorithm, fuzzy logic and related work techniques are discussed with their advantages. GA is an optimization algorithm that can help in finding appropriate fuzzy rules and fuzzy rule is a machine learning algorithm. Fuzzy- genetic based approach provides performance better than GA based techniques [21]. Due to increasing incidents of attacks on network securing data is a prime goal. More hybrid techniques should be investigated in this area. In misuse detection mode signature of new intrusions should be created so that it is easy to catch the attack. This paper will prove a good starting point for newcomers in the field of GA and fuzzy logic based intrusion detection and is useful for people looking for a quick review of recent development in this field.

ACKNOWLEDGMENT

The heading of the Acknowledgment section and the References section must not be numbered. Causal Productions wishes to acknowledge Michael Shell and other contributors for developing and maintaining the IEEE LaTeX style files which have been used in the preparation of this template.

REFERENCES

[1] Harjinder Kaur and N. Gill, "Host based anomaly detection using fuzzy genetic approach (FGA)," International Journal of Computer Applications, vol. 74, (2013).

[2] Mohammad Sazzadul Hoque, Md. Abdul Mukit, and M. A. N. Bikas, "An implementation of intrusion detection system using genetic algorithm," International Journal of Network Security & Its Applications (IJNSA), vol. 4, (2012).

[3] Mostaque Md and M. Hassan, "Current studies on intrusion detection system, genetic and fuzzy logic," International Journal of Distributed and Parallel Systems (IJDPS), vol. 4, (2013).

[4] Kuldeep Kumar and R. Punia, "Improving the performance of IDS using Genetic Algorithm," International Journal of Computer Science & Communication, vol. 4, pp. (93-98),(2013).

[5] A. Einipour, "Intelligent Intrusion Detection in Computer Networks Using Fuzzy Systems," Global Journal of Computer Science and Technology Neural & Artificial Intelligence, vol. 12, (2012).

[6] UCI, "Task," 1999. Available online at: http://kdd.ics.uci.edu/databases/kddcup99/task.html. Last Accessed: 3rd April, (2014).

[7] M. Abliz, "Internet Denial of Service Attacks and Defense Mechanisms," University of Pittsburgh, Pittsburgh, Technical Report TR-11-178,(2011).
[8] Rashid Husain and S. Muhammad, "A survey on soft computing techniques in network security," Scholarly Journal of Mathematics and Computer Science vol. 2, pp. (28-32), (2013).
[9] Susan M. Bridges and R. B. Vaughn, "Fuzzy data mining and genetic algorithms applied to intrusion detection," in 23rd National Information Security Conference, (2000).
[10] S. N. Pawar, "Intrusion detection in computer network using genetic algorithm approach: a survey," International Journal of Advances in Engineering & Technology, (2013).
[11] K. S. Anil Kumar and V. N. Mohan. (2012). Adaptive genetic algorithm model for intrusion detection. Available online at: http://core.kmi.open.ac.uk/download/pdf/9329382.pdf. Last Accessed: 21st February, (2014).
[12] B. Uppalaiah, Anand Narsimha Swaraj, and T. Bharat, "Genetic algorithm approach to intrusion detection system," International Journal of Computer Science and Telecommunications, vol. 3, (2012).
[13] Shiyong Li, Yanli Zhu, Lijuan Ma, and Y. Liang, "Research on intrusion detection based on genetic algorithm and rough set," in International Conference on Electrical and Computer Engineering Advances in Biomedical Engineering, (2012).
[14] Kamal Kishore Prasad and S. Borah, "Use of genetic algorithms in intrusion detection systems: an analysis," International Journal of Applied Research and Studies, vol. 2, (2013).
[15] Shaokun Liu, Lina Yu, and Y. Fang, "Application of improved genetic algorithm in reliability optimization of multi-agent intrusion detection," in 2nd International Conference on Electronic & Mechanical Engineering and Information Technology,(2012).
[16] Madhuri Agravat and U. P. Rao, "Computer intrusion detection by two objective fuzzy genetic algorithm," in CCSEA 2011, pp. (281–292),(2011).
[17] P. A. Diaz-Gomez, "Improved off-line intrusion detection using a genetic algorithm," in ICIES 2005, pp. (66-73), (2005).
[18] P. Jongsuebsuk, N. Wattanapongsakorn, and C. Charnsripinyo, "Real-time intrusion detection with fuzzy genetic algorithm," in 10th International Conference on Electrical Engineering/Electronics, Computer, Telecommunication and Information Technology (ECTI-CON), pp. (1-6), (2013).
[19] N. Wattanapongsakorn, S. Srakaew E. Wonghirunsombat, C. Sribavonmongkol, T. Junhom, and P. Jongsubsook, "A practical network-based intrusion detection and prevention system," in IEEE 11th International Conference on Trust, Security and Privacy in Computing and Communications, (2012).
[20] Y. Wang, "Using Fuzzy Expert System Based on Genetic Algorithms for Intrusion Detection System," in International Forum on Information Technology and Applications, pp. (221-224), (2009).
[21] T. P. Fries, "A fuzzy-genetic approach to network intrusion detection," in GECCO'08 Conference Companion on Genetic and Evolutionary Computation, Atlanta, Georgia, USA, pp. (2141-2146),(2008).
[22] Hee-su Chae, Byung-oh Jo, Sang-Hyun Choi and Twae-kyung Park, "Feature Selection for Intrusion Detection using NSL-KDD" Proceeding of the 6th WSEAS World Congress Applied Computing Conference (2013).
[23] Iftikhar Ahmad, Azween B Abdullah and Abdullah S Alghamdi, "Application of Artificial Neural Network in Detection of Probing Attacks" IEEE Symposium on Industrial Electronics and Applications (2009).
[24] Vitaly Shmatikov and Ming-Hsiu Wang, "Security Against Probe-Response Attacks in Collaborative Intrusion Detection" Proceeding of the 2007 Workshop on Large Scale Attack Defense ACM (2007

BIOGRAPHY

Authors have the option to publish a biography together with the paper, with the academic qualification, past and present positions, research interests, awards, etc. This increases the profile of the authors and is well received by international reader.