ISSN (Online) 2581-9429

IJARST

International Journal of Advanced Research in Science & Technology (IJARST)

Volume 8, Issue 1, August 2020

Reducing Attributes Using Data Mining Techniques and Genetic Algorithm for the Development of Intrusion Detection System

Miss Priyanka R. Pidurkar[1] and Dr. R. R. Keole[2]
M.E. Scholar, Computer Science and Engineering[1]
Associate Professor and Head, Department of Information Technology[2]
Dr. Rajendra Gode Institute of Technology and Research, Amravati, India[1]
Hanuman Vyayam Prasarak Mandal's College of Engineering and Technology, Amravati, India[2]
priyankapidurkar6@gmail.com[1] and ranjitkeole@gmail.com[2]

Abstract: *As the higher attribute of systems allows greater exposure to outside people and makes it possible for attackers to avoid recognition, the drawback of detection is a difficult task as network protocols can mutate. Intrusion monitoring systems are used to prevent abnormal device access. I will be providing this paper with a review of the techniques of genetic rule intrusion detection. The IDS, which is described as a system security response, is commonly used to detect suspicious activities over a device or network. For intrusion identifications, altogether separate methods are used at this stage, but sadly, none of the methods is ideal. Hunting for new technologies therefore continues. As I progressed, I developed an AN Intrusion Detection System (IDS) through the application of the genetic rule (GA) to rapidly detect various intrusive network activities. The intrusion detection data collection NSL-KDD is used to analyze and test the proposed intrusion identification scheme. The test results reveal precisely that, regardless of how much the documents are conventional or unusual or not, the organized systems obtained a better accurate rate and acquired a relatively inexpensive rate of identification.*

Keywords: Intrusion Detection System(IDS), Genetic Algorithm, Data mining, NSL-KDD dataset.

I. Introduction

The intruder is an infringement or invasion of the limitations. And it is an unwelcome contribution to this study. This is an unlawful means not to enter or take possession of the property of the owner or to enter, seize, or take possession of the property of the owner through unethical channels. The entry made before the remaining person or the reversionary after perseverance of a specific succession is incorrect. The sometimes illegal act of hacking, snooping, and stealing data by cyber-space is an event involving illegal access to the data or an automated communications network. Victims that are ignorant of their vulnerability become cyber-intrusions. The data may be threatened or in danger without security measures and controls in all parts of the world. Some of the attacks are timid, so the details can already be tracked. On the other hand, certain attacks allow information to be altered intentionally and the data or the network itself to be modified or lost. Intrusion detection is being developed into the required technologies to track the traffic of the network and detect network infringements, such as irregular network behavior, unlawful network entry, and suspicious assaults on portable computer systems[2].

The main components of the fuzzy logic (FL), neural artificial networks (ANNs), probabilistic reasoning (PR), and genetic algorithms are the sentimental computing sector. The soft computing methods region unit usually used with rules-based skilled systems to obtain specific information when we use detection capabilities anywhere and everywhere data is diagrammatic as a set fuzzy rule-then rule. Given the preparation of radically new soft computing methods, the likelihood of maltreatment is often underused for the intrusion detection systems. In this study, we aim to address a network abuse detection method focused on genetic algorithms. GA was chosen for a variety of its nice features, e.g., resilient to noise, no gradient knowledge is needed to check for an optimal or under-optimal global response, self-learning features, etc. Any network assaults the integrity, privacy, and accessibility of a system. We want intrusion prevention systems to reduce such an impact on a channel. There are various units of intrusion detection devices,

DOI: XX.082020/IJARST

including primarily host IDS, mostly network-based IDS. The host mainly runs IDS on the system individually. Regarding unauthorized behavior, the network primarily monitors network IDS traffic. Intrusion information for attack type-Intrusion is loose to four different forms of attack[4][6], as mentioned below: Intrusion information.

1. The popular service called, Denial of service might be a group in which any partner perpetrator is very active or too complete to manage legitimate demands for a computer code segment and/or memory space, thereby refusing reasonable needs network access.
2. Unapproved remote-computer access can be a type in which a partner attacker transfers the message over a network to a system and then utilizes the weakness of the system to illegally obtain native accessibility as a user.
3. Unexpected root access performs a type of attack on the field where the attacker attempts to use the weakness to reach a root system from a standard user profile on the network.
4. The Inquisitor is a type of attack whether a hacker searches a network to gather information or execute better-known vulnerabilities. The Inquisitor is the type of attack. The information can be accessed by an individual with a list of computers and facilities that are situated in a network.
5. One type of scanning attack involves sending the data packet to the network, trying to collect topological information, open or closed ports, kinds of traffic that are allowed, active hosts on a network, or which kinds of software programs or types are running. Some of these scanning methodologies may be employed by delusional SQL injection attacks when trying to find more vulnerable points on a network. Attacks are also scanned for open ports that can inject malware or malicious code.

A. IDS (Intrusion Detection System) Placement

The location of intrusion prevention devices relies on the network and is important. This method gives IDS good traffic visibility into the network and doesn't take traffic from network users. The edge of the network is where an extranet can be linked. A technique whereby a technician places his first IDS at the point of maximum exposure and the other at the next highest point depending on the resources accessible will proceed until all areas of the networks are reached. Another realistic solution which can be done if more assets are accessible. When mounted over a network firewall, an IDS can mainly be used to shield itself from internet traffic, but more specifically to protect it from specific threats, such as port scanners and network mapmakers. In this place, IDS will track and be signature-based layers 4-7 of the OSI model. It is highly useful because there will be attempted breaches to minimize the number of false positives rather than revealing real network compromises that have occurred through the firewall. The IDS in this role also helps to reduce the time taken to detect effective network threats. All that in effect decreases costs and operating uncertainty. The real network also provides another way to put IDS. They show threats or unusual network behavior. Ignoring network protection can create a great many issues, causing users to lose safety or encouraging an intruder to walk about easily that has broken into the network. Strenuous intranet protection makes maneuvering and improving access impossible for only certain users inside the network.

B. Scope of the Generated Framework

A computer or software application with any or all these specific functions is an intrusion detection system (IDS):

1. Monitors a whole cyber threat network infrastructure
2. Detects a malware threat as it occurs automatically
3. Deploys easily countermeasures (intruder defense systems) to avoid the attack
4. Submit information to a network administrator or team

IDS is geared at developing an integrated system for cyber threat detection and involving a team of live security researchers who will respond to the intrusion attempt. They will then display the activity 's data review and then apply measures to enhance network security. The intrusion prevention program is used to secure any aspect of the networks in on-site network infrastructure, a virtual server or a cloud-based platform, including software, hardware, and software. It constitutes a physical boundary that safeguards the IT network of business partly or entirely.

II. Literature Review

A. Basic Terminologies

The preceding two types of IDS can be classified [3]:

Signature Based IDS: IDS packages dependent on signatures on the network, along with a collection of signatures or characteristics from documented acts of malice. The way most virus-based software identifies ransomware is close. The concern is that a difference would occur between a new hazard that has been identified and an IDS signature[4].

Anomaly-Based IDS: The anomaly-based IDS tracks and contrasts network activity to a given benchmark. Within this base layer, the network follows the same form of bandwidth, protocols used, ports and hardware communication, and where the traffic detections that is abnormal, and substantially different are observed, alerts the controller or customer to the baseline. IDS is used in several ways to track unusual flow.

a. IDS (System of Intrusion Detection) Basic Issues

IDS is not an automated defense alternative: IDS could not necessarily deliver 100% threat detection security. Instead, it operates with a firewall, antivirus, and other security controls. Consequently, an efficient system to use IDS in a wider security policy. IDS can't be scaled: IDS could not deal with further development or business growth because of its efficiency restrictions. Throughput issues will certainly arise when businesses adopt web-based services or hybrid networks comprising both on-site and web applications. The more businesses are moving physical processes into the cloud, the more IDS would prove insufficient. False Positives & False Negatives: Because of its critical control capability, IDS will give a false warning, even though there is no indication of the operation, if the device detects a security breach. Even then, one more important problem is that, because of the vast volume of data traffic, IDS sensors frequently fail to identify suspicious attacks. So many false positives and/or negatives generating devices pose various problems for managers with budget limitations who use reliable data to reduce security risks.

Administrators Need to be Experienced: To operate an IDS, an skilled IT professional has to learn and expertise. Personnel who lack the experience and expertise to run an IDS can need to be more involved in network management. Inexperienced workers may also react to an assault more gradually. Therefore, the network cannot be properly secured.

Packets Encrypted: An IDS cannot be able to track packets that are encrypted. This unlocks a doorway that enables attackers to enter an unrecognized network. The intruder could have influenced the network to control confidential data as soon as the IDS identifies the data leak.

Attacks depending on the Protocol: Also caught protocols can be analyzed by an IDS. It is also vulnerable as a network host to the very same protocol threats. The IDS could malfunction glitches as well as corrupted reports.

b. Genetic Algorithm

This is an optimization approach derived from natural adaptation and biology evolutionary principles. It is evolutionary. GAs is an imaginative use of a random check to address problems more efficiently. While randomized, GAs may not use random information to guide the query in a search field with greater results, instead[6]. Rather they use historical data. A vector representation exercise feature was developed by Howard H. Rosenbrock in 1960[7] as a statistical optimization technique. GA uses vectorized objective functions for smoother processing and has an optimal performance. The GA only calls fitness once but assumes that all entities in the existing population can measure fitness at once.[8][9] This is GA's maximum frequency.

c. Methodologies of Evasion

There are various methods employed by terrorists, which are called 'fast' steps to be avoided. The intruder would be under the scanner by transmitting broken packages and quickly circumvent the functionality of the tracking sensor to monitor the signature of the threat.

1. The protocol-used TCP port does not necessarily signify the transferred protocol. The IDS would not be able to identify the existence of a trojan if the intruder reorganized it to use a different port.
2. It is complicated for IDS to compare captured packets and to conclude that there is a network scan in development by organizing search among multiple attackers and assigning specific ports or hosts for specific

DOI: XX.082020/IJARST

perpetrators.

3. The security manager's ability to identify the origins of the threat will increase the complexity by using poorly managed or improperly installed proxy servers to recover from the attack. It is very difficult for IDS to identify the source of the problem if the source is impersonated and bounces by a browser.
4. IDS usually focuses on 'pattern matching' for threat detection. It could be necessary to prevent identification by modifying the data utilized in the threat slightly.

B. Work-Related

Any of the methods for intrusion identification are summarized described here. Nevertheless, in the last section of the paper, some GA-based IDSs are addressed to equate the research with our research and to contrast it. During intrusion identification systems, many studies have attempted GA differently. For weighted function abstraction with a particular framework for intrusion prevention results, Melani J Middlemiss et al. (2003) used GA. Our genetic algorithms have been simple and weight evolve for the characteristics of the data collection. [18] For the GA fitness function and for evaluating the efficiency of the new randomized unit package, the k-nearest neighbor classifier was utilized. The test was successful and the exercise was accurate[18].

The methodology for implementing GA in IDSs is developed by Wei Li (2004). He addressed various application specifics after a short presentation on IDS, the GA, and associated detection technologies. GA has used the guidelines for classifying regular network connections from abnormal connections[10] to produce classification laws. A basic genetic algorithm has been used by RenHui Gong et al. (2005)[20] to derivate from system auditing data a series of classification rules. The Help Confidence System is employed as fitness to determine each law consistency [20]. The new feature selection approach is introduced in Chi Hoon Lee et al. (2006), maximizing the class distinction between regular computer network interfaces and attack patterns. The genetic algorithm was used by Saqib Ashfaq et al. (2006) for the development of appropriate guidelines for cost-sensitive harassment identification of intruders. They also used M.J.Middlemiss et al 's five most weighted characteristics. To define such attributes, they built a GA. The algorithm creates rules, such that the correct action could be taken, which will define a threat as well as its type. The cost-sensitive method takes into consideration the risk of false alerts individually for each form of threat. Nalini N. And Rao G. Raghavendra. (2006) introduce a modern system of genetic algorithms intrusion detection and the key factor analysis[19]. Blessyrajra and Dr. A.J.Deepa (2016) have suggested a method " Enhanced Detection Guard System against malware in Network " An improved EDGS is employed to authenticate users in the managed links in this analysis. To identify the families of malware, EDGS uses a heuristic approach to evaluating breaches. Heuristic detection can identify many previously known ransomware and new malware variants. Entropy Metric and J-Metric are used in the heuristic algorithm. The characteristics selection is critical to boosting data extraction algorithms' efficiency[7]. Various scholars suggest various algorithms, from Bayesian methods [8] to decision-making trees [9], from law model [10] to learning functions [11], in different groups. Thus the measurement improvements are improving and increasing as well. Scientists have focused on heuristic and hyperheuristic approaches to function selection in recent times. Some examples include the Genetic Algorithm[12], Particle Swarm Optimization[13], and Ant Colony Optimization[14] are the most common methods. The well- known SVM-dependent FS closed-loop system, called SVM-RFE, was introduced by Sung and Mukkamala, who extracted one function at a period sequentially and evaluated output for each SVM test[15]. Six main characteristics were also graded [16].

Author	Techniques	Summary	Dataset Utilized
M. Ghalehgolabi and A. Rezaeipanah in 2016 [16]	Genetic algorithm optimization	This paper proposed a framework for the development of IDS by identifying primary attributes and that system to be efficient and effective in terms of computations.	NSL-KDD
S. Lakhina et al. in 2010 [17]	Principal Component Analysis and neural	The paper proposed a framework with the development of a hybridized model with the combination of principal	NSL-KDD

	network	component analysis and neural network. PCA helps reduce the no. of attributes while neural networks identify the new threats depending on the training.	
G. K. Kuchimanchi et al. in 2004 [18]	Feature Extraction, Principal component analysis, and Neural Network	The proposed framework developed using feature extraction and neural network with 2 different approaches such as NNPCA, and NLCA. The Gain in Information used as a metric for feature extraction.	KDD Cup 99
J. Kratica et al. in 2012 [20]	Feature Selection, Genetic Algorithm	The proposed approach to implement and justify a methodology for solving the problem such as RCSP and also presented various comparisons of other heuristic techniques with the present technique	
Y. Xu et al.in 2017 [21]	Support Vector Machine, Information Gain, swarm optimization	The proposed framework for the Particle identification of malware with the help of statistical techniques such as support vector machine, information gain for the computation of weights of the various features, and particle swarm optimization for the method optimization.	Faker91, FakeInst, SkullKey, Aiplay, DroidDream, https://virusshare.com
S. Y. Yerima et al. in 2015 [22]	Bayesian Classification	This document mainly dealt with mobile malware especially the case of android. The development of this framework based on the machine learning technique i.e., Bayesian classification through static analysis.	Real-time data using certain android apk applications
W. Wang et al. in 2009 [23]	Support vector machine, KNN, Naïve-Bayes Classification, CART, Random forest	This paper proposed a framework to identify mal-apps and categorizing those benign apps by using various techniques such as SVM, KNN, Naïve-Bayes, CART, RF classifications though ensembling.	Anzi [24] VirusTotal[25]

Table I: Summary of the Work-Related Papers Considered

III. Methodology

A. Proposed System

There are other attributes of other data sets like NSL-KDD. In the classification of results, all these attributes do not, on the other hand, play a positive role. You would then pick the best apps from a subset. A genetic algorithm is used in this work to pick the desired characteristics. This approach is focused on the study of the functions of the distribution system. This element contributes to boosting the chromosomes of genetic algorithms by understanding their particularities. A dataset in various measurements can be used for the proposed procedure. The popular techniques of data mining such as the decision tree and the K-nearest neighbor are utilized for the evaluations of the chosen functions. The flowchart of the suggested approach as seen in Figure 1.

a. Preprocessing Data

Pre-processing of the data is the first stage in the development of any model dependent on data mining methodologies. Pre-processing is undertaken to optimize and enhance the accuracy of actual data for processing. This move includes the translation, normalization, and disassembly of data from string to list.

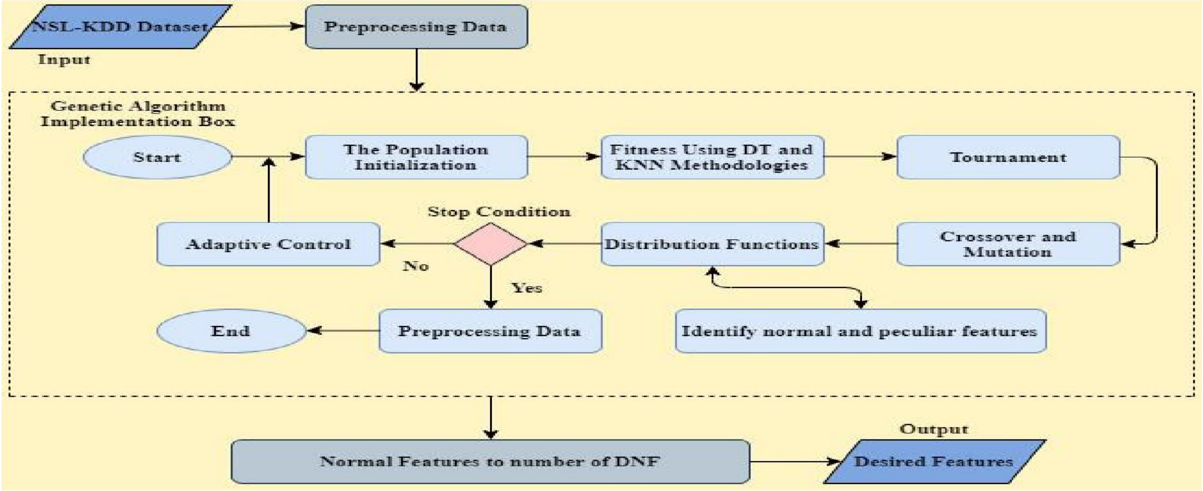

Figure 1: Intrusion Detection System (IDS) Proposed System Architecture

b. Function Quality Improved Genetic Algorithms

The genetic algorithm, influenced by biology and Darwinian evolutionary theories, was published by Holland in 1970. In this analysis, the chromosome arrangement of the number of each attribute is considered. Every chromosome is a string of 0 and 1 bits with a cumulative number of characteristics each. The chromosome genes illustrate the beneficial properties that are used in the data processing. The number of required features (DNFs) is set for checking and malfunction in this investigation. The suggested approach sequentially applies the genetic algorithm, which allows the primary population of features to be generated per repeat. The genetic algorithm continues at variance with the original chromosome population. Then the puzzled and uncertain characteristics of the quest field are collected and used in later stages to produce the population. Compact characteristics are a set of characteristics used in population growth. Specialty is a vector of characteristics that are not suitable for pre-population usage or used in the new population's growth. The chromosome fitness test is the failure rate in sample classification. Two KNN and DT classifiers were used owing to the fast fitness measurement.

The development of transitional-generation chromosomes with high fitness is one of the important phenomenons of the genetic algorithm. These chromosomes can be killed and can no longer be formed as a result of the implementation of mutant and crossover workers. An allele with the highest fitness rate is automatically transferred to the next generation of each family.

IV. Implementation

A. Algorithm Discussion

The flowchart of the methodology through proposed system architecture as shown in Fig. 1. The implementation can be classified into various stages such as initialization, evaluating fitness, selection, crossover, and, finally, mutation.

a. Implemented Algorithm

1. The population needs to be initialized.
2. The number of records in the NSL-KDD dataset considered to be N.
3. In the dataset, for each chromosome
4. Initialize A with 0, and AB with 0

DOI: XX.082020/IJARST

```
5.  For every record in the dataset
6.  If the record can be comparable with the chromosome
7.  then AB should be incremented by 1 (AB + 1).
8.  End if
9.  If the record can only be compared with the “condition” section
10. then A should be incremented by 1 (A + 1).
11. End if
12. End for
13. Evaluate the Fitness value by using 1 / (1+F_obj), and F_obj can be evaluated as ((a + 2b + 3c + 4d) - 41)
14. If the evaluated Fitness of chromosome > among all the evaluated values of the Fitness
15. Choose that particular chromosome, and add into the new population
16. End if
17. End for
18. For every chromosome that added into the new population
19. crossover operator should be applied to the chromosome
20. mutation operator should be applied to the chromosome
21. End for
22. If the number of generations still exists, then move to step 4.
```

B. Discussion on Processing Steps

a. Stage-01:

During the first stage a zero matrix will be formed or a matrix with "0" will be simply formed in any position, which means that we may assume that it is a random start. Then you have to pick from one of the attributes in the NSL KDD data collection, which means that we will choose every column from 41 attributes and fill out the column row by row or column with a null matrix. But we must be mindful that the function column has to be arbitrarily chosen from our initial NSL-KDD dataset.

b. Stage-02:

The fitness function of each chromosome is primarily discussed at this point. At this point, we will measure each chromosome or gene 's fitness function using this formulation.

$$\text{Fitness} = 1 / (1+f\_obj) \qquad (1)$$

Where f_obj represents the function of (a + 2b + 3c + 4d) - 41)

c. Stage-03:

It was the first fitness assessment incarnation. There only two chromosomal fitness values are chosen in the highest among all genes. For more measures, chromosomes are chosen for over 93.7 percent fitness and 96 percent fitness value. Now we need to do this step over and over again by increasing the fitness value for each new chromosome and then selecting the highest value or genes or chromosomes for more stages, indicating the same process in iteration 2 and 3.

d. Stage-04:

There are two types of genetic algorithms, namely if we want to have more and more precision for fitness values:

Crossover: conjugation (crossover) is the very first phase in the replication process. Through it, a brand new chromosome is created by the chromosomes of the parents. The traditional recombination of the GA is an action that requires two parents, but which also includes strategies with more parent areas. The traditional (distributed) crossover and blended (transitional) crossover are two of the most prominent utilized algorithms. Within this segment, two of the most fitness worth genes have to be taken and crossover operations added to them to then generate new chromosomes.

Upon crossbreeding, there have been two new generations of genes or chromosomes, the fitness value of these newly created genes is now to be determined and the two highest chromosomes are compared to previous fitness values. We now need the two genes with the highest fitness value to pick. We have to select two genes for further estimation here with maximum value. The measurement must add a mutation surgery to better fitness.

Mutation: The mutation occurs essentially by modifying the value of the position and then measuring its fitness value. The newly developed population of selection and fusion can be used for mutation further. Mutation means that row or gene components are modified. Such changes can be caused by errors during the parent gene copy procedure. So far as GA is concerned, the mutation is a random variation of the population origin of the gene. The chromosome that is the cell and the cell itself is also randomly picked. Now we'll have to pick between the two highest fitness-rated chromosomes from 4 genes after completing the transition. We need to choose only 2 genes. Of various matrix configurations of NSL-KDD Datasets, we need to run the same procedures and evaluate the objective function of better precision.

V. Discussions About Results

The procedure has to be applied or carried out even more for extreme precision of performance measurements in an intrusion detection system dependent on the genetic algorithm. We, therefore, need to run all our MATLAB tests using NSL-KDD as a dataset. NSL-KDD DATASET Specific statistical studies that influenced the precise estimates of many IDS modeled by researchers have shown the intrinsic inconvenience in the KDD cup 99 datasets. The dataset NSL-KDD is its previous administration's polished version. The full KDD data set includes important records. The researchers are supplied with a set of downloadable files.

A. Evaluation Metrics

For measuring or the identification of efficiency of a framework generated statistically, we considered certain evaluation metrics such as accuracy, sensitivity, specificity, and these can be calculated as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$Sensitivity\ (TPR) = \frac{TP}{TP + FN}$$

$$Specificity\ (TNR) = \frac{TN}{FP + TN}$$

Where TP is a true positive, TN is a true negative, FP is a false positive, FN is a false negative respectively.

S. No.	Utilized Dataset	Methodology	Evaluation Metrics	
			Accuracy	Specificity
1	KDD CUP 99 Dataset	EDGS uses a Heuristic algorithm	0.87 0.92	0.74 0.95
2	NSL-KDD Dataset	Genetic algorithm	0.92 0.93	0.96 0.97

Table II: Summary of the Results Comparison

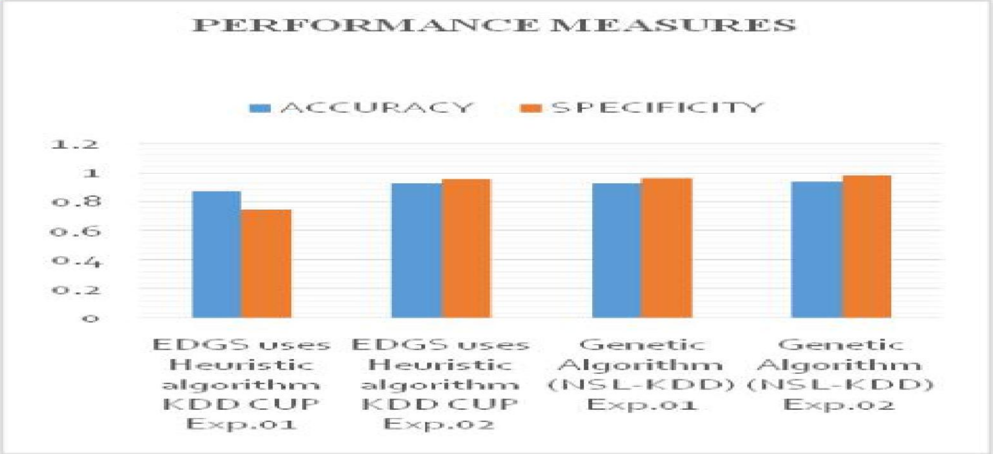

Figure 4: Comparison of the evaluation metrics of the result

The key difference is the number of specifics between the NSL-KDD and KDD 99. The number of rows about 65,536 and the number of columns about 41 in the case of the KDD 99 dataset. However, NSL-KDD has fewer rows of data. The 25,193 rows NSL-KDD dataset and 41.sets may conclude that the same precision with the less information is obtained by the use of the NSL-KDD dataset. It's an optimization technique as well. So is often used with the NSL-KDD data set to measure a certain precision and the same precision with less information.

B. Advantages

In case of conflict or threat, the network or device is tracked indefinitely. Based on customer needs the system may be updated and adjusted and can lead to the external and internal systems and network risks. It avoids any network disruption effectively. It offers an easy-to-use GUI for simple information management systems. Changes to machine files and folders can be identified and recorded quickly. In-network packages you should adapt to the different content: Firewalls can show the ports and IP addresses used between the two hosts.

C. Disadvantages

The intrusion detection system's drawback is that the threat vector can not be identified and in any case, the entire network is locked. You won't prevent accidents by yourself: an IDS doesn't stop or prevent threats, it helps to expose them. As a result, an IDS needs to be included in the overall plan that includes additional safety measures and staff that know how to respond appropriately. Data packet will still be misleading: IDS reading of the IP packet details can always be spoofed with network addresses. When an intruder uses fake details, this makes it harder to identify and analyze the threat. False Positives Another critical flaw with IDS is often that they warn you repeatedly to false positives. False positives are in so many ways more common than actual threats. The number of false positives could be selected to minimize an IDS but the engineers would still need the ability to reply. Unless the false positives are not tracked, actual attacks will slip away or be missed.

VI. Conclusion and Future Scope

A genetic algorithm that detects various kinds of network intrusions and fraud identification is used to present and enforce an IDS. Implementing neural networks ensures that an intrusion detection system model can be used in the future to make more precision and use the NSL-KDD dataset standard because this data set provides a higher precision through the use of less preference functionality. This work also concerns the results by function decreased of many intrusion detection systems. Simply put, the optimization procedure may be performed. Intruders will also see increased use of target tracking in the long run in the intended detection. As already stated, target tracking has been one of the most consistent and effective intrusion detection techniques. Intruders make changes to the processes almost inevitably, sometimes to build backdoors, but alterations often happen just by mistake (particularly for inexperienced intruders. Intruders could be able to avoid signature IDs and also erase records of device logs to hide evidence of their operation.

But, they are less likely to abandon the attention of a target tracking method that utilizes a range of unique authenticated, encryption, and decryption protocols for connectivity to target tracking. Although commercial target tracking systems like Tripwire and Intruder Warning are commonly utilized in Fortune 500 firms, the expense of using them in other applications also dissuades their utilization in small firms. Through using the standard NSL- KDD data collection we measure the efficiency of our system and obtain an appropriate detection limit. In short, we can seek to improve our infringement recognition scheme by employing two frameworks or can in combination with any framework based on or technique of data mining say IDS based on genetic algorithm.

REFERENCES

[1]. R.Elamaran and R.Mala, “A Study on Network Intrusion Detection System (NIDS) In Virtual Network Structure”, International Journal of Computer Sciences and Engineering (IJCSE), Vol. 03, Issue - 11, November-2015, pp.59 – 164.

[2]. Hassan, Mostaque Md. "Current studies on intrusion detection system, genetic algorithm and fuzzy logic." arXiv preprint arXiv:1304.3535 (2013).

[3]. Dhanalakshmi, Y., and I. Ramesh Babu. "Intrusion detection using data mining along fuzzy logic and genetic algorithms." International Journal of ComputerScience and Network Security 8.2 (2008): 27-32.

[4]. Zorana Banković, José M. Moya, Álvaro Araujo, Slobodan Bojanić and Octavio Nieto-Taladriz, “A Genetic Algorithm based Solution for Intrusion Detection”, Journal of Information Assurance and Security(JIAS), Vol.4, Issue-3, June-2009, pp.192-199

[5]. Shelly Xiaonan Wu, Wolfgang Banzhaf, “The use of computational intelligence in intrusion detection systems: a review”, Applied Soft Computing, Vol.10, Issue-01, January-2010, pp.1–35.

[6]. Mohammad Sazzadul Hoque, Md. Abdul Mukit & Md. Abu Naser Bikas, “An Implementation of Intrusion Detection System using Genetic Algorithm”, Department of Computer Science and Engineering, Shahjalal University of Science and Technology, Sylhet, Bangladesh, International Journal of Network Security and Its Applications (IJNSA), Vol.4, Issue-2, March-2012, pp.109-120.

[7]. S Selvakani Kandeeban, and Rengan S Rajesh, Department of Computer Applications, Jaya Engineering College1 Chennai, Tamilnadu, 602 024, India. Dept. of CSE, MS University, Tirunelveli, Tamilnadu, 627 009, India, “Integrated Intrusion Detection System using Soft Computing”, International Journal of Network Security, Vol.10, Issue-2, March -2010,pp.87-92.

[8]. W. Lu and I. Traore, Department of Electrical and Computer Engineering, University of Victoria, Victoria B.C., Canada “Detecting New Forms of Network Intrusion Using Genetic Programming”, Computational Intelligence, vol. 20, Issue-03, August -2004, pp.475-494.

[9]. K. Burbeck & N.Y. Simmin (2007), Department of Computer and Information Science, Linkoping University, Sweden, “Adaptive

[10]. T.S. Chou, K.K. Yen & J. Luo, “Network Intrusion Detection Design using Feature Selection of Soft Computing Paradigms”, International Journal of Computational Intelligence, Vol. 4, Issue-3, 2008, pp.196–208.

[11]. Khan, Latifur, Mamoun Awad, and Bhavani Thuraisingham. "A new intrusion detection system using support vector machines and hierarchical clustering." The VLDB journal 16.4 (2007): 507-521.

[12]. Burney, SM Aqil, M. Sadiq Ali Khan, and Mr Jawed Naseem. "Efficient Probabilistic Classification Methods for NIDS." International Journal of ComputerScience & Information Security (2010).

[13]. Wang, B., Li, F., & Zhang, S. (2009, November). Research on intrusion detection based on campus network. In 2009 Third International Symposium onIntelligent Information Technology Application (Vol. 1, pp. 468-471). IEEE.

[14]. L.Dhanabal, Dr. S.P. Shantharajah, Assistant Professor [SRG], Dept. of Computer Applications, Kumaraguru College of Technology, Coimbatore, India, Professor, Department of MCA, Sona College of Technology, Salem, India, “A Study on NSLKDD Dataset for Intrusion Detection System Based on Classification

DOI: XX.082020/IJARST

Algorithms", International Journal of Advanced Research in Computer and Communication Engineering, Vol. 4, Issue 6, June 2015, pp.446-452.

[15]. R.P. Purushottam, Dr. Yogesh Sharma, and Dr. Manali Kshirsagar, "using Genetic Algorithm : A Study," Int. J. Emerg. Technol. Comput. Sci., vol. 3, no. 2, pp. 282–286, 2014.

[16]. M. Ghalehgolabi and A. Rezaeipanah, "Intrusion Detection System Using Genetic Algorithm and Data Mining Techniques Based on the Reduction Features," Int. J. Comput. Appl. Technol. Res., vol. 6, no. 11, pp. 461–466, 2016.

[17]. S. Lakhina, S. Joseph, and B. Verma, "Feature Reduction using Principal Component Analysis for Effective Anomaly–Based Intrusion Detection on NSL-KDD," Int. J. Eng. Sci. Technol., vol. 2, no. 6, pp. 1790–1799, 2010.

[18]. G. K. Kuchimanchi, V. V. Phoha, K. S. Balagani and S. R. Gaddam, "Dimension reduction using feature extraction methods for real-time misuse detection systems," Proceedings from the Fifth Annual IEEE SMC Information Assurance Workshop, 2004., West Point, NY, 2004, pp. 195-202, doi: 10.1109/IAW.2004.1437817.

[19]. E. M. Karabulut, S. A. Özel, and T. İbrikçi, "A comparative study on the effect of feature selection on classification accuracy," Procedia Technol., vol. 1, pp. 323–327, 2012.

[20]. J. Kratica, T. Kostić, D. Tǒsić, D. Dugǒsija, and V. Filipović, "A genetic algorithm for the routing and carrier selection problem," Comput. Sci. Inf. Syst., vol. 9, no. 1, pp. 49–62, 2012.

[21]. Y. Xu, C. Wu, K. Zheng, X. Wang, X. Niu, and T. Lu, "Computing Adaptive Feature Weights with PSO to Improve Android Malware Detection," Secur.Commun. Networks, vol. 2017, no. May, 2017.

[22]. S. Y. Yerima, S. Sezer, and G. McWilliams, "Analysis of Bayesian classification-based approaches for Android malware detection," IET Inf. Secur., vol. 8, no. 1, pp. 25–36, 2014.

[23]. W. Wang, Y. Li, X. Wang, J. Liu, and X. Zhang, "Detecting Android malicious apps and categorizing benign apps with ensemble of classifiers," Futur.Gener. Comput. Syst., vol. 78, pp. 987–994, 2018.

[24]. Anzhi, http://www.anzhi.com/

[25]. Virus Total , https://www.virustotal.com/