Ensemble Framework of Artificial immune system based on Network Intrusion Detection System for Network Security Sustainability

*Sarangam Kodati [1 *], Nara Sreekanth [2] , K.S.R.K.Sarma [3], P Chandra Sekhar Reddy [4], Archana Saxena [5], Boya Palajonna Narasaiah [6]*

[1]Department of Information Technology, CVR College of Engineering, Hyderabad, Telangana, India
[2]Department of CSE, BVRIT HYDERABAD College for Women, Hyderabad, Telangana, India
[3]CSE Department, Vidya Jyothi Institute of Technology, Hyderabad, Telangana, India
[4]Department, Gokaraju Rangaraju Institute of Engineering and Technology, Hyderabad, Telangana, India
[5]Uttaranchal Institute of Management, Uttaranchal University, Dehradun, India
[6]KG Reddy College of Engineering & Technology, Hyderabad, Telangana, India

Abstract: The popularity and rapid growth of the internet have reemphasized the importance of intrusion detection systems (IDS) significance in the network security. IDS decreases hacking, data theft risk, privacy intrusion, and others. To save the system from external and internal intruders, the primary approaches of IDS are used. Many techniques[13], like genetic algorithms, artificial neural networks, and artificial immune systems, have been applied to IDS. This paper describes an Ensemble Framework of Artificial Immune System (AIS) based on Network Intrusion Detection System. Without placing a significant additional load on networks and monitoring systems, the large volume of data is analysed by a network-based Intrusion Detection System (NIDS). For determining the connection type, data from KDD Cup 99 competitions is utilized. To differentiate between attacks and valid connections, IDS can be utilized. Optimized feature selection is used to speed up the time-consuming rough set. The results obtained from the IDS system indicate that it can effectively identify the attacking connections with a high success rate.

1 Introduction

A security measurement system that is used to protect the system from malicious attacks that can compromise the integrity and confidentiality of resource information is an IDS [1]. For internet usage, the foremost prerequisite is secure data exchange, if usage increases rapidly, then network vulnerabilities also increase [2]. Due to the attacks, the network systems are jeopardize continuously and threatens data confidentiality and seamless service usage. Various models of IDS are suggested and developed for protecting the system from attacks [3]. These models have the ability to detect, in particular, and prevent malicious attacks over the network, retaining the system's normal performance and integrity during malicious attacks.

With each increment of device connectivity, the number of attacks on the device increases, which in turn increases the need for network security [4]. Society depends more on connected devices due to the advent of interconnected cars, power technology, and hospitals. And the network faces real threats to the physical safety of users [5]. Monitoring the network traffic for attack detection and prevention is one of the techniques for improving network security. Damages will be controlled and reversed, possibly by the early detection of attacks. The commercially available IDS systems often use rules that can define attacks, like antivirus programs. The intrusion detection systems were using signatures that were manually listed within the community used in the system (e.g., SNORT). Complex programs increasing and networks made this task hard, steadily increasing the size of the signature list.

*Corresponding author: k.sarangam@gmail.com

Basically, the intrusion detection system is mainly classified into strategies. One is anomaly detection, then another is misuse detection [6–7].From normal patterns, the anomaly detection system analyses the divergence, being watchful for novel attacks or unknown behaviour without having any knowledge about it [8]. Anomaly detection systems have the ability to detect unrecognized attacks, which can cause incorrect alarms at a higher percentage. A manual detection system relies on data from previous intrusion patterns. It is a actual efficient then accurate method for the detection of an intrusion. But its drawback is its inability to detect the novel attack. The misuse detection technique uses known patterns of various attacks to identify the known attacks [9].

To identify the misuse and anomalies of the system, various machine learning techniques have been introduced [10]. The internet plays a significant role in the daily activities of humans, having unlimited storage space for data storage. The device that is related to the internet is unprotected to hacking, and computer security can be violated. This requires key attention for securing data inside the system. In this situation, the IDS comes into play to picture for protecting user information from any type of intrusion. Different techniques have been used for intrusion development, like Neural networks, SVM (Support Vector Machines), and GA (Genetic algorithms) [11–12] and other methods[13-18]. For specific types of attacks, these techniques have a good rate of detection, unable to identify other types of attacks. The main goal of this system is to develop an intelligent intrusion system for enhancing the system's performance. The immune system is used to improve accuracy. To enhance the system's performance, reduce the intrusion detection system's complexity. The experiment is conducted in two phases: training and testing. Antibodies (Ab) initial solutions are generated during the training stage. The antibodies pool contains six various types of attack connections: Neptune, Portsweep, Smurf, Satan, Ipsweep, and Land.

2 Artificial Intrusion Detection

2.1 Immune systems

During the research of various algorithms, a look at nature gives feasible solutions as a structure of the immune system. By using the immune system, the host body is protected from unwanted antigens (for example, proteins on the foreign virus surface) without responding to the body itself. A complex system that has been extensively studied is the HIS (human immune system), where silica full reproduction is still not possible and system parts have been replicated already. Most importantly, the clonal selection and negative selection processes have been identified for the merit of intrusion detection systems. Many research projects have been going on inspired by the immune system, algorithms; intrusion detection systems real-world applications are yet an open field.

2.1.1 The Human Immune System

There are two types of human immune systems: the NIS (natural immune system) and the AIS (adaptive immune system). In NIS "constructed in" the cells are already existing inside the human body and protect against the predefined pathogens. It includes molecules, organs, and cells. The NIS is likewise accountable for protecting the human body from foreign, risky pathogens. The adaptive immune system produces antibodies for protective the human frame in opposition to pathogens, it's far primarily based on self-sample recognition. it may include lymphocytes. The lymphocytes consist of more antibodies (B-cellular and T-cell) and are accountable for the detection and rejection of foreign pathogens.

2.1.2 Artificial Immune system

For implementing computer security, most of the research attempts to understand the meaning of human immunity using immune system mathematical modeling and multilevel defense to protect the system against intruder behaviour. There are two types of AIS: adaptive immune systems and innate immune systems. The immune system provides the first line of defense against intrusion patterns. It can be made to match between the stored intrusion system and host pattern depending on the past pattern intruder's definition. The adaptive immune system produces antibodies to defend the system against intruders. Based on self-pattern recognition, the AIS made a match between the knowledge base of self-recognition and the host pattern.

IDS contains hardware and software devices; during intrusion detection, it makes alarms. The IDS is mainly categorized into two kinds: host-based and network-based IDS. In the host-based system, the sensor contains some application-based and software agent IDS; it analyses all the host activities where it is installed and includes logs, file systems, and the kernel. In NIDS, to capture and monitor all the network traffic, the sensors are located at choke points along the network border.

3 ‘AIS’ Based on Network Intrusion Detection System Framework

The process of the IDS system is represented as a flow chart, as shown in Fig. 1. The process began with generating antibodies in the initial population, which were executed randomly.

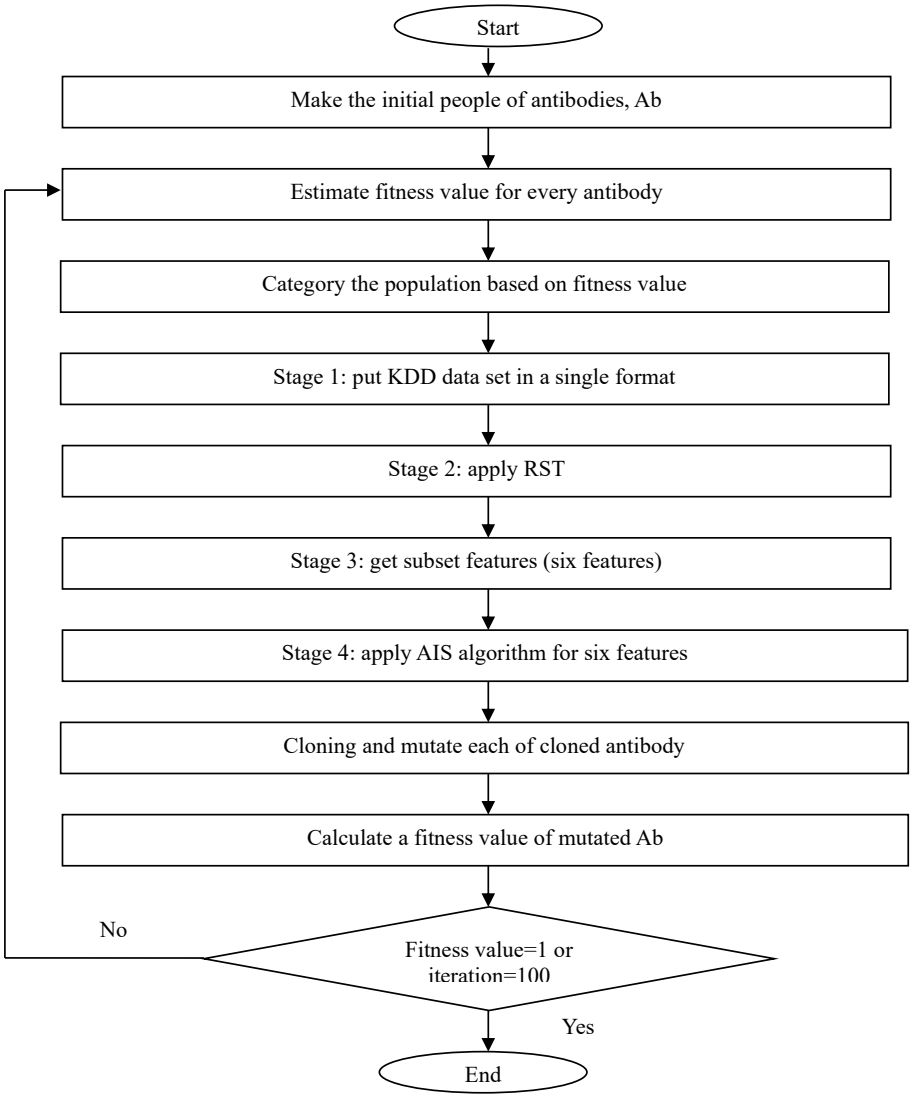

Fig. 1. 'AIS' Based on the Network Intrusion Detection System Framework

Through utilized a specific objective function, each antibody fitness value was calculated. In this step, the number of attacks patterns that every antibody can identify is determined. If the antibody can identify a huge quantity of patterns, then it will have high fitness. Based on fitness value, the antibody population can be sorted. The top ten antibodies with higher fitness values are involved in the next step. To produce a new antibody collection, the selected antibodies are changed and cloned.

3.1 Experimental Data Sets

For detecting the network intrusion, the standard KDDcup99 dataset is utilized here. Many researchers have employed this dataset in this environment. Original datasets contain millions of information associates, which represent normal as well as attack. Only selected attack connections are investigated due to the research size and time constraints. Every connection has 24 features that can differentiate between an attack and a normal scenario. The main dataset contains more than 2,00,000 connections. The datasets are separated into 2 groups: training and testing. In the dataset, every

recorded connection has 41 features, which include one class label. The label specifies whether the connection is normal or an attack. In this paper, only 24 distinctive features are investigated among 41 features. The reason behind this is that some of the features are different from attack of normal connections. So the irrelevant features are removed from this approach. Here four kinds of attacks are investigated:

1)Dos: DOS means Denial of Service. The DOS attack rejects the legitimate requests and takes more memory resources, computing. DOS can vary from buffer overflows to system resource flooding. Examples are teardrop, syn flood, smurf, and ping-of-death.

2) U2R: U2R means user to root**.** It can provide unauthorized access to the root privileges of a user. This type of attack enters the system as a common user, and vulnerabilities in the system can be exploited gradually to attain superuser access. Examples of this type of attack are buffer overflow attack.

3) Probe: The attacker is scanning the network and finding known vulnerabilities in the system to gather network information. To attack the system, known vulnerabilities can be exploited. An example of this type of attack is port scanning.

4) R2L: R2L means remote to local. From remote machine to local machine, unauthorized access is provided by these. An example of this type of attack is guessing passwords.

3.2 Fitness Value Calculation

Each generated antibody's quality can be evaluated using an objective function. The antibody's fitness can be determined by this function. The representation of the objective function as a mathematical calculation 1 is below:

$$F = \frac{a}{A} - \frac{b}{B} \qquad (1)$$

The above equation representing that generated antibodies control to are expecting the variety of attack connections effectively. Here, the number of attacks in a dataset is represented by A, whereas b is the antibody that can correctly predict the quantity of ordinary connections from the total normal connections presented in dataset B. The values of A and B are constant, rather than a and b, which depend on the high-quality of the generated antibody. The received health value lies between -1 and 1 in this process. A fitness value near 1 denotes the ability of the generated body to predict the attack connections for datasets in a number of the hassle instances. Hence, in the direction of 1 health value approach a higher antibody is generated.

3.3 Principles of the immune system

Conversions of the fundamentals of HIS into mathematical tools by so many researchers are used for many applications, like artificial immune systems based on NIDS. Immune system principles and their applications in the framework of intrusion detection systems can be introduced.

• **Multilevel protection:** The HIS affords a multilevel protection system like outside protection and different members of the mucosal body's. For detecting an attack, the IDS system contains multilevel defenses like a firewall as external defenses. Suppose the first level of IDS fails to identify the attack, then the second level of IDS would be enabled, and so on.

• **Innate immunity:** The HIS contains molecules and "built-in" cells. The initial response to any kind of pathogen is the responsibility of this system. IDS consists of public resources and numerous databases for providing expert knowledge of attacks on computers.

4 Experimental Results

In this paper, the KDDcup99 data set is used, it has normal and no-flood connection vectors, rough set theory, and an artificial immune system. The KDDcup99 dataset has 284948 data connections. Among these, 10% of the data connections, i.e., 28494, are selected for testing. Based on the predetermined probability, the selection process was randomly executed. For the training process, the remaining data connections, i.e., 256454, were used. A probability of 0.2 is used to initiate the experimental process. The KDDcup99 dataset has a 20% possibility per attack connection, which is being chosen for testing. This step is carried out one after the other for each information connection within the dataset until all 28494 connections have been used. Rough set theory is applied to the test data for eliminating redundant or irrelevant features in the dataset. This theory of rough set provides a decision table with fewer attribute

features; the reduction result is 6 attributes (flag, rv_count, Service, dst_host_srv Jatein, src_bytes). The AIS algorithm is finally applied to the KDDcup99 dataset, which has six attributes.

The final population contains antibodies that go through the process of evaluation through mutation and cloning (genetic) operations. So, high-quality antibodies are expected. During the testing process to test the data earlier, selected datasets are used. For examining the test data, high-quality antibodies are used. The TPR (True-positive rate) can be calculated depending on the amount of testing datasets in which the antibodies are predicted correctly. The TPR rate can be calculated by using the following equation 2.

$$\text{TPR} = \frac{\text{number of attack correctly predicted}}{\text{total number of attacks}} \text{ X } 100 \qquad (2)$$

To know the effect of using various probabilities for the selection process in this paper, experiments with another set of probabilities were conducted. The investigated probability values are 0.3, 0.4, and 0.5. These values and experimental results are shown below in Table 1.

Table 1. Different Probability Values 'TPR'

Probability	**Number of attacks correctly predefined**	**TPR (%)**
0.3	27654	97.12
0.4	28053	98.34
0.5	28432	99.67

Summarization of Table 1 is that the AIS recognizes more attack connections as probability increases. The 0.3 probability value provides a TPR of 97.12%, TPR 0.4 probability value produces a success rate of 98.34% for the algorithm. The better result was provided by a 0.5 probability value with a 99.67% success rate. A probability value of 0.5 denotes that every information connection in the dataset has 50% possibilities of being selected for the method of checking out. As in advance said, only 10% of the initial dataset connections are selected for testing. So the selection process was completed one by one from the beginning dataset; if high selection value probabilities were used, then it was assumed that the dataset connections were selected from the earlier part.

The aim of this training process is to provide antibodies with higher fitness values. For recognizing the attacks that are simulated during the process of training, the important tools are good quality attributes. As earlier said, the fitness values lie between -1 and 1. Fitness value 1 refers to good quality antibody.

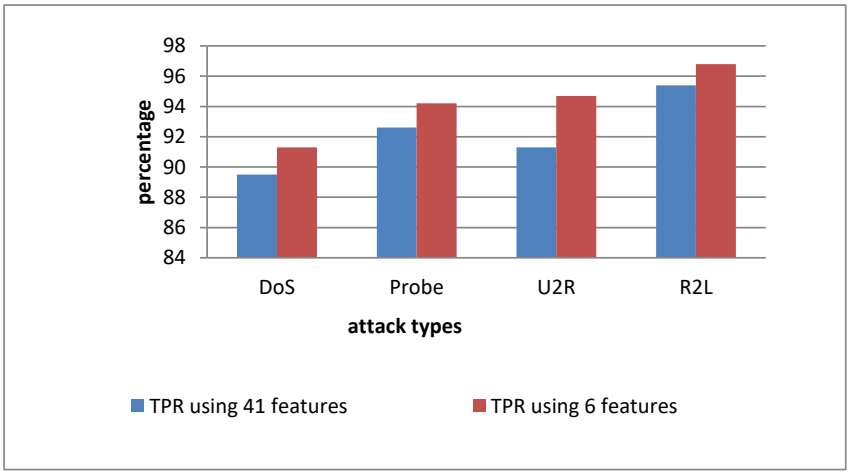

Fig. 2. Different Attacks True Positive Rate

After applying the AIS to the KDDcup99 dataset, having 41 features gives high TPR for every type of attack, but TPR will be reduced using rough set theory. From Fig. 2, it is observed that the TPR and TNR (true negative rate) detection become highly relative.

5 Conclusion

The intrusion detection system's efficient performance achieves high accuracy and takes less time for detection. Here, An Ensemble Framework of Artificial immune systems (AIS) based on NIDS was suggested. The primary aspects of this system are the mutation and cloning evaluation processes for producing good quality antibodies in the testing process. For investigating the AIS's robustness, the KDDcup99 dataset was used. The success rate, or TPR, provided by the AIS is discussed, for addressing the problem. The effectiveness of the AIS is highlighted. Cooperation between cells, generalization, and multilevel defense are provided by AIS. The complexity problem of the KDDcup99 dataset is solved by rough set theory. The RST reduces the KDDcup99 data set features from 41 to 6. The TNR and TPR became high.

References

1. AshwiniKatkar, SakshiShukla, Danish Shaikh, PradipDange, *Malware Intrusion Detection For System Security*, International Conference on Communication information and Computing Technology (ICCICT), (2021)
2. A Abdul Rasheed, *Vulnerability detection towards protecting intrusion by Social Network Analysis approach*, 5th International Conference on Trends in Electronics and Informatics (ICOEI), (2021)
3. SibiAmaran, R. Madhan Mohan, *Intrusion Detection System using Optimal Support Vector Machine for Wireless Sensor Networks*, International Conference on Artificial Intelligence and Smart Systems (ICAIS), (2021)
4. Yaofu Cao, Xiaomeng Li, Shulin Zhang, Yang Li, Liang Chen, Yunrui He, *Design of network security situation awareness analysis module for electric power dispatching and control system*, 2nd International Conference on Information Technology and Computer Application (ITCA), (2020)
5. S. Prabhu, Mary Anita E.A., *Trust based secure routing mechanisms for wireless sensor networks: A survey*, 6th International Conference on Advanced Computing and Communication Systems (ICACCS), (2020)
6. Zhaoyue Wu, Hongjun Su, Qian Du, *Low-Rank and Collaborative Representation for Hyperspectral Anomaly Detection*, IGARSS 2019 - IEEE International Geoscience and Remote Sensing Symposium, (2019)
7. Dimitrios Papamartzivanos, Félix Gómez Mármol, GeorgiosKambourakis, IEEE Access. **7**, (2019)
8. P. V. Haripriya, J. S. Anju, *An AIS based anomaly detection system*, International Conference on Computing Methodologies and Communication (ICCMC), (2017)
9. Rohini Rajpal, Sanmeet Kaur, Raman deep Kaur, *Improving detection rate using misuse detection and machine learning*, SAI Computing Conference (SAI), (2016)
10. Tahir Mehmood, Helmi B. MdRais, *Machine learning algorithms in context of intrusion detection*, 3rd International Conference on Computer and Information Sciences (ICCOINS), (2016)
11. Praneeth Nskh, M Naveen Varma, Roshan Ramakrishna Naik, Principle component analysis based intrusion detection system using support vector machine", IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT), (2016)
12. Yogita Danane, Thaksen Parvat, "Intrusion detection system using fuzzy genetic algorithm", International Conference on Pervasive Computing (ICPC), (2015)
13. Nagaraja, A., Boregowda, U., Khatatneh, K., Vangipuram, R., Nuvvusetty, R., Sravan Kiran, V., IEEE Access, 8, art. no. 9006824, pp. 39184-39196. (2020)
14. Pradeep, G., Sakthidharan, G.R, *A Survey on Performance Comparison of Support Vector Machine, Random Forest, and Extreme Learning Machine for Intrusion Detection*, International Conference On Machine Learning Big Data Management Cloud and Computing, (2021)
15. Sri Vidya, M., Sakthidharan, G.R., *Accurate Anomaly Detection using various Machine Learning methods for IoT devices in Indoor Environment*, Proceedings of the 5th International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud), I-SMAC (2021)
16. Shalli Rani, M. Balasaraswathi, P. Chandra Sekhar Reddy, Gurbinder Singh Brar, M. Sivaram & Vigneswaran Dhasarathan, Wireless Networks, **26**, (2020)
17. P Chandra Sekhar Reddy, G. R. Sakthidharan, International Journal of Intelligent Engineering and Systems, **12**, (2019)
18. N. Madhusudhana Reddy, G. Ramesh, Srinivasa Babu Kasturi, D. Sharmila, G. Gopichand & L. Thomas Robinson, Applied Nanoscience, **13**, (2023).