Hong Kong International Journal of Research Studies, ISSN: 3078-4018
Volume 1, Issue 2, July December, 2023
Available online at: https://octopuspublication.com/index.php/hkijrs

For Preventative Protection of Cloud Infrastructure, An Intrusion Detection System Powered By AI Was Implemented.

Suresh S[1], Dr. Manisha[2]

[1]Research Scholar, Department of Computer Science & Engineering, JS University, Shikohabad, UP
[2]Assistant Professor, Supervisor, Department of Computer Science & Engineering, JS University, Shikohabad, UP

ABSTRACT

Among the many modern intrusion detection systems available, the Intrusion Detection System (IDS) stands out. Key goals of this program include protecting sensitive data, adding another level of protection, and preventing unauthorized access to networks. Intrusion detection systems (IDS) scan all network traffic for malicious content and alert administrators to any suspicious activity, protecting hosts and networks from harm. Also, alarm systems are programmed to detect out-of-the-ordinary behavior. Because of the rapid expansion of the Internet, many new industries have emerged. Big data, cloud computing, and the IoT are all instances of these emerging markets. One possible explanation for the uptick in attack frequency might be the dramatic rise in data creation and transmission rates throughout the network.

This is why many academics have concentrated on intrusion detection systems (IDS) and helped make them better at warding off assaults and other dangers related to these types of attacks. Nonetheless, a large portion of the information included in network logs probably contains characteristics that are unimportant for attack detection or categorization. Thus, experts still struggle to learn anything useful from this kind of network data and determine whether the selected attributes may improve the efficiency of immune system detection systems (IDS). Furthermore, for breach detection systems to handle the wide variety of threats, a large collection is required. Determining the primary features is a tough but crucial step in improving the intrusion detection system's (IDS) speed and accuracy.

Various machine learning, deep learning, and evolutionary algorithms are used by the present intrusion detection system (IDS) to identify threats and learn from historical data based on trends.

Because these solutions consider every facet of traffic simultaneously, it is probable that they will be expensive to execute. However, these approaches yield desirable results. As a result, the issue of reducing costs without sacrificing efficiency or adding features that aren't essential must be continuously addressed. Nevertheless, FSA analysis of the NSL-KDD and CICIDS2017 files was conducted first in attempt to resolve these concerns.

This was done so that we could focus on the most important traits while removing the ones that weren't necessary. It was essential to develop more cost-effective intrusion detection systems (IDS) that could function in very vast networks in order to meet this demand. In order to improve the intrusion detection system's (IDS) detection engine, we look at and assess several models that employ FSA with NSL-KDD datasets.

Keywords: Intrusion Detection System (IDS), Feature Selection Algorithm (FSA),Machine Learning, Network Security

INTRODUCTION

The use of internet services has been very commonplace in today's globe throughout the last several decades. All hours of the day and night, from any location, most clients utilize these services on a wide range of electronic devices such smartphones, laptops, tablets, and more. This means that these networks might be used to send sensitive or vital data. The continual transfer of private data between devices and data centers for archival and retrieval reasons is another effect of the ever-developing internet.

There is a window of opportunity for the attackers to launch several attacks that can endanger the targeted target because of these consequences. Potentially, an attacker may use a variety of state-of-the-art techniques to exploit system security holes. The system might be compromised if unauthorized users get access to it, which could lead to the disclosure of sensitive

information or the breach of their accounts. Contemporary security solutions are crucial for shielding system administrators and security personnel from the hazards of today. New technology such as the Internet of Things and big data are contributing to the ever-increasing data flow.

The result is an increase in data congestion on the network, making it slower, more difficult, and more difficult to change the assault profile. Another critical skill for data scientists, companies, and marketers is the ability to sift through massive data sets for actionable insights. Academics and scientists worried about network security are starting to notice the sheer amount of data generated by these connections. The ever-increasing user base of the internet is the primary reason for this. Network security refers to the study of taking measures to prevent unauthorized persons from gaining access to computer systems or networks by detecting and fixing security vulnerabilities. There have been several defensive solutions developed over the last twenty years to ward against threats including denial-of-service (DoS), user-to-root (U2R), remote-to-local (R2L), probing, and many more.

Firewalls and antivirus software are only two examples. This is why fundamental security mechanisms must be in place to detect new forms of attacks and malicious data or traffic that might compromise the system or network. An intrusion detection system (IDS) is a name for this kind of tool [1]. Common parlance calls them "IDSs." An intrusion detection system (IDS) gathers, processes, and identifies incoming data using a combination of software and hardware technologies. By using these tools, threats such as fraudulent attacks, potential dangers, and annoying network and individual systems may be located and eliminated [2]. An intrusion detection system (IDS) is responsible for protecting sensitive data as it is sent over a network. To accomplish these tasks and reach these objectives, it is required to examine the intrusion detection system (IDS), analyze its data using mathematical or statistical methods, and make sure that it alerts network administrators and managers of any suspicious activity [3].

A. AN OUTLINE OF THE ISSUE

With the advent of COVID-19 and other internet-based threats, the importance of having access to these services has skyrocketed during the last 20 years. The quick ascent to fame could be explained by the advent of much better Internet technology. In order to access these services quickly and from anywhere, people commonly utilize electronic devices like computers, tablets, cellphones, and similar ones. This means that more and more sensitive data is being moved between computers and data storage facilities over these networks. This opens the door for thieves to launch massive assaults that might compromise the company or its customers by evading security measures. In their pursuit of security holes in computer systems, attackers use a plethora of intricate tactics. Several of these approaches are detailed in this article. This might lead to the exploitation of sensitive data, the theft of user accounts, or the acquisition of unauthorized access to the system. To lessen the impact of these attacks, professionals and researchers are focusing on protecting important information and fortifying networks.

The broad usage of intrusion detection systems, also referred to as IDS, has provided a solution. Intrusion detection systems examine data as it is entered to determine whether it pertains to system-wide or network-wide activities. The quantity of data being created and exchanged inside the network has increased due to the spread of technologies like the internet, social media, and the Internet of Things. This is a result of how popular these tools have become. Network traffic might potentially have a wide range of side effects, some of which could be rather bothersome and others quite insignificant. To tackle this issue, successful intrusion detection systems (IDS) will provide several monitoring methodologies in addition to ways for adding or removing features. Its significance in halting the expansion of the system's processing power and working time cannot be emphasized enough. For this reason, models for feature reduction or deletion and fast decision-making engines were created to address this issue [17]. It is likely not the most effective approach to compare and assess several models using only one classifier or estimate.

2. BACKGROUND

The proliferation of new network types (such as software-defined and wireless sensor networks) and the dynamic nature of attack tactics is making it harder than ever to keep computer systems secure. Prioritizing security measures was not a top priority during the initial setup of these newer networks. In most cases, these networks are not adequately protected by the conventional security measures. Having a rapid security system that can detect attempts to breach a computer system is crucial in light of this fact. The requirement for an intrusion detection system led to the development of what is now commonly known as an intrusion detection system (IDS). A primary function of intrusion detection systems (IDS) is to monitor networks for intrusions [19]. In order to make sure that authenticity, integrity, and confidentiality—the three cornerstones of computer security—are still intact, this is done. It is done with the aim of finding any potential dangers or infractions.

Hong Kong International Journal of Research Studies, ISSN: 3078-4018
Volume 1, Issue 2, July December, 2023
Available online at: https://octopuspublication.com/index.php/hkijrs

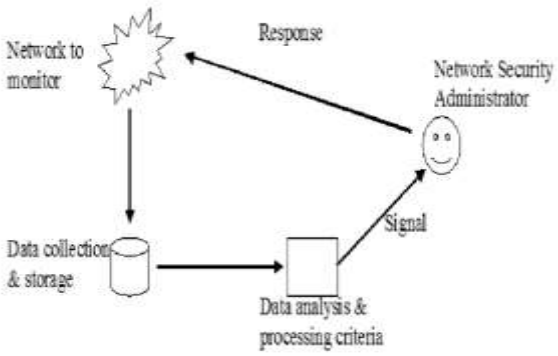


Figure 2.2 General IDS's architecture.

In figure 2.2, we are able to see the usual design for intrusion detection systems (IDS) that was created by Axelsson [25]. "IDS components" is an umbrella word that encompasses a wide variety of elements, including the following:

Monitoring a network on a consistent basis is required in order to identify any vulnerabilities that may exist inside the network. Take into consideration the following as evidence:

It is possible that this might be used for either a single computer or a network, depending on the circumstances.

2. Following the collection of data from a variety of sources, the data storage and retrieval unit organizes the data in an appropriate manner and saves it on a disk in a secure manner. The data processing and analysis unit used by an intrusion detection system (IDS) functions as the system's central nervous system. It is possible for the system to identify attack flow patterns that might potentially cause damage.

4. Signal The network master is supplied with a warning signal that contains information on the targeted assault. This component of the system is responsible for being in charge of monitoring and managing the output of the Intrusion Detection System (IDS). Either it will fix the problem or it will alert the administrator of the network to the next actions they need to take.It is stated in the notion that the person who is accountable for the network security system should react promptly upon receiving a warning signal that indicates the discovery of a threat.

It is possible that the conclusion will be a preset reaction to an incursion or a signal to the administrator of the network security system of possibly dangerous activities.

An intrusion detection system (IDS) is designed to monitor, investigate, and react to any suspicious behavior or policy violations that may occur inside a computer network or host system. This is the fundamental aim of an IDS. The purpose of an intrusion detection system, often known as an IDS, is to monitor all of the system's data and activity in order to discover any unusual patterns that may signal security weaknesses, abuse, or attacks. Intrusion detection systems (IDS) may have different architectures and approaches to threat detection, but in general, they all include a number of components that are similar to one another and work together to identify threats.

A. CLASSIFICATION OF IDS

After giving serious thought to the matter, the intrusion detection system (IDS) arranges the information sources and locations that it monitors. Various intrusion detection systems were developed as a result of the discovery of anomalies, the implementation of detection procedures, and the consideration of security issues.

It might be response infrastructure, response instruments, or response systems.
It is vital to classify intrusion detection systems (IDS) in order to have an understanding of their functioning, deployment, and role in recognizing and preventing security risks. For the purpose of identifying any indications of malicious activity, policy breaches, or odd behavior, intrusion detection systems perform continuous monitoring of systems and networks. Professionals in the security industry are able to provide answers to every issue by categorizing items into different groups. The design, deployment, threat detection capabilities, data analysis time, and response strategy of intrusion detection systems (IDSs) may

provide the basis for categorizing these systems into several groups. Each of the categories adds to comprehensive security monitoring by emphasizing a different characteristic of the intrusion detection system (IDS).

When it comes to IDS classification, one of the most crucial aspects is precisely determining the deployment site. Different hosts or servers are responsible for monitoring various aspects of the system, including system calls, file security, user logins, configuration changes, application logs, and host-based intrusion detection systems (HIDS). Hidden threats, unlawful access, and malicious behavior that does not damage publicly accessible network data may be identified by having a high-level intrusion detection system (HIDS). On the other hand, managing HIDS across several sites may be difficult and take a significant amount of resources. Network-Based Intrusion Detection Systems (NIDS), on the other hand, are able to monitor communication between devices by strategically locating themselves inside a network. By examining the contents and headers of packets, network intrusion detection systems (NIDS) have the ability to identify threats to networks. These threats include port scanning, the propagation of malware, and denial-of-service assaults. Although network intrusion detection systems (NIDS) have the capability to protect several systems at the same time, their efficiency may be undermined if critical data is encrypted or if a single server is the only destination of an assault. The purpose of mixed intrusion detection systems (IDS) is to integrate host-based and network-based methodologies in order to improve the sensitivity and accuracy of their detections.

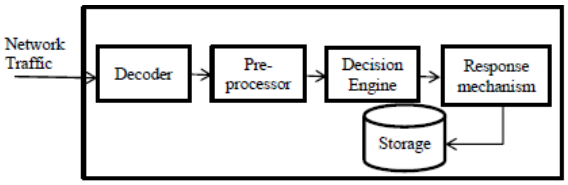


Figure 2.5: Architecture of traditional NIDS

B.IDS'STAXONOMY

As its name suggests, an intrusion detection system (IDS) is able to quickly identify and notify any odd behavior by monitoring a network or other system. This is accomplished by maintaining a close check on the system. Liao et al. [31] provide a classification of intrusion detection systems that is based on four fundamental characteristics. Availability of information, instability of the system, detection method, and response time are all factors to consider." Figure 2.4 is an illustration of the different Intrusion Detection Systems (IDSs) that are stated that can be found in reference [31].

The classification of intrusion detection systems (IDSs) enables us to arrange and grasp IDS technology in accordance with its application, attack identification capabilities, data source, response methods, and system architecture. This category may be used by specialists, system designers, and security managers in order to choose the intrusion detection system (IDS) solutions that are the most appropriate for the company's needs, network size, threat models, and performance requirements. The vocabulary that is used to describe intrusion detection systems has also changed in tandem with the growing sophistication of cyber threats. The consequence of this is the development of detection methods that are both more complex and multi-modalities.

One of the most important aspects of IDS classification is the categorization of intrusion detection systems (IDSs) according to their deployment location, which in turn shows the system component that the IDS is a part of. Host-based intrusion detection systems, also known as HIDS, are responsible for monitoring system calls, application logs, changes to file systems, and user activity on each individual host or server location.

Through the use of HIDS, it is possible to identify attempts at insider threats, elevated privileges, and unwanted file alterations all at the same time. Additionally, they provide easy access to a wide variety of local activities. On the other hand, they need to be installed and maintained on every system that is being monitored, which may result in an increase in costs in settings that are on a big scale. On the other hand, network intrusion detection systems (NIDSs) are strategically positioned at crucial network nodes such ports, switches, and routers in order to continually monitor all network traffic.

The contents and headers of packets are analyzed by network intrusion detection systems (NIDS) in order to identify vulnerabilities that are network-based, assaults that denial of service, and odd communication patterns. It is possible for network intrusion detection systems (NIDS) to struggle with protected data and fail to offer a full picture of host-level activities, despite the fact that they have the capability to defend several systems at the same time. The integration of host-based and network-based methodologies is what hybrid intrusion detection systems (IDS) do in order to give complete and comprehensive protection. This is accomplished by increasing the beneficial qualities of each approach while simultaneously minimizing the bad aspects of each method.

Dimensionality reduction remains an unsolved major problem in data science and information research. Over the past few decades, a plethora of IDS models have been developed to aid in the process of dimensionality reduction in very big datasets. Here you may find all of the networks, including KDD99 and NSL KDD. In order to improve intrusion detection systems (IDS) and avoid problems associated with high-dimensional data, the FSA has been used in several research. However, professionals still have challenges when it comes to simplifying data and managing tasks [43]. The number of potential dangers has grown in tandem with the exponential growth of network traffic. Since this is the case, several specialists have used various FSA-based machine learning techniques for Intrusion Detection Systems (IDS).

The evaluation of intrusion detection systems was accomplished by Mukkamala and colleagues [44] with the help of Support Vector Machines (SVM) and neural networks (NN). Throughout the testing, Support Vector Machines (SVMs) proved to be very efficient and versatile when dealing with big datasets. Learning will need a tremendous investment of time from NN. To determine which attributes were relevant to this discussion, Fleuret et al. (2004) used a method called the joint information methodology. When combined with a Bayes network, this approach outperforms SVM alone. Their research has mostly focused on overall work time [45]. Intrusion detection systems (IDS) were the subject of study by Chebrolu and colleagues in 2005. Bayes networks, reverse classification trees, and other state-of-the-art technical breakthroughs were used in the inquiry. They successfully detected and evaded several assault kinds by making use of twelve critical traits derived from their method. Surprisingly high detection rates of U2R assaults have been shown [46]. Problems related to multi-dimensional data were addressed by Chou et al. (2008) using a variety of novel feature selection algorithms (FSAs), including correlation-based feature selection (CFS) and fast CFS. One issue is that the data keep repeating themselves and are not specific enough.

4. A MODEL FOR INTRUSION DETECTION SYSTEMS BASED ON AN EMBEDDED LEARNING ALGORITHM

This chapter lays the groundwork for FST-based systems by identifying key components that may improve the recognition engine's performance. Recursive feature elimination (RFE) is a method that has been used by a wide range of algorithms to improve their performance.

Completed with the aim and all necessary attributes in tow, the target has been reached. We utilize the NSL KDD Dataset for method development and evaluation. You can demonstrate the difference in accuracy between choosing any features and selecting the right features by doing this. Here we compare and examine the performance of RFE with other ensemble classifiers, such as GB, AB, ET, and RF classifiers. Additional classifiers are also taken into account. Sorting the data into several categories is what these algorithms are all about. In order to greatly enhance the classifier's success rate and overall performance, comparative analysis found that carefully selecting the necessary attributes was the key. Here is a synopsis of the primary resources that I used when writing this chapter.

To determine the significance of the components, the RFE method was used in conjunction with the ANOVA F-Test and the select_Percentile procedure after the UFS had been run.

A. COLLECTION DATA

In the last two decades, researchers have made extensive use of the KDD 1999 dataset, which was first developed at Lincoln Labs at the Massachusetts Institute of Technology [89]. As a result of the enhancements that were developed for the NSL-KDD dataset, the KDD1999 dataset has become even more superior (81). There are a lot of problems that will be solved by this compilation.

It is possible to explain the KDD 1999 dataset in the following manner:
It is possible that the objective outputs of our algorithms, which are developed from a variety of different methodologies, may result in a higher recognition rate or accuracy on regular data. The lack of duplicate data in both the training set and the testing set creates the possibility that this will be possible. inside the original KDD 1999 dataset, there are two unique parts that include the total of the items (records) that are contained inside the training dataset and the testing dataset. This eliminates the possibility of any particular piece of information being repeated.

There are a number of persuasive reasons in favor of utilizing the NSL-KDD dataset, including the following: It is important to eliminate similar data in order to make the algorithm produce more objective conclusions. A significant number of events may be found in both the training dataset and the testing dataset. It is possible that tests may be conducted on the whole collection rather than choosing pieces at random from a smaller and smaller group. As a last point of interest,

Hong Kong International Journal of Research Studies, ISSN: 3078-4018
Volume 1, Issue 2, July December, 2023
Available online at: https://octopuspublication.com/index.php/hkijrs

it offers a multitude of features, including exact network architecture, thorough packet capture, well-organized notes, and a multitude of other benefits.

Features Types	Description	Examples
Basic_Features	TCP/IP is the source of the features.	Services, flag, duration land, urgent etc.
Content Features	To gain access to the initial TCP's payload. These features use domain knowledge.	Num_root, Num_shell, hot, Logged_in etc.
Time-based traffic Features	Consider features that span two-second temporal frames have been captured.	Rerror_rate, Srv_count, count, Serror_rate etc.
Host-based traffic Features	These features have the same destination host as the current connections are accessed and span greater than two seconds intervals.	Dist_hst_srv_count, Dist_hst_same_srv_rate etc.

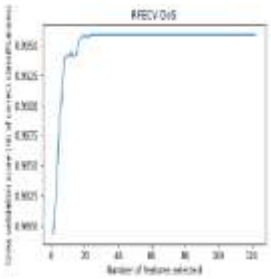


Figure 4.3 DoS RFECV with AB

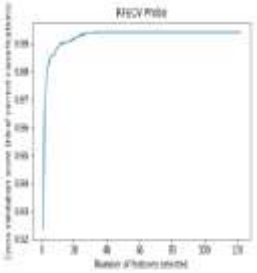


Figure 4.4 PROBE RFECV with AB

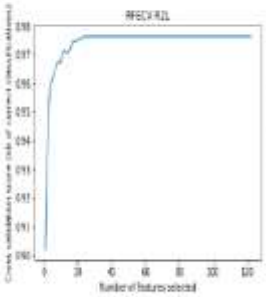


Figure 4.5 R2L RFECV with AB

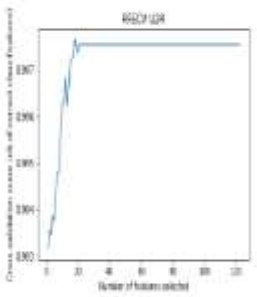


Figure 4.6 U2R RFECV with AB

5. ADAPTIVE IDS FRAMEWORK.

In order to help you identify crucial features and remove irrelevant ones, this chapter will show you a new way to use FSA. On top of that, it verifies that the NSL-KDD datasets are really intrusion detection system (IDS) sensitive. In order to determine the best accurate predictor, the engine was subjected to a battery of tests. It was possible to evaluate the best FST and predictor using the CICIDS2017 real-time dataset. In this chapter, we see examples of testing that show how the key features of the proposed model improve IDS performance while drastically reducing processing requirements. The accuracy was raised by 99.21% on the NSL-KDD dataset and by 99.94% on the CICIDS2017 dataset when using the suggested model. Testing was the means by which both of these outcomes were achieved.

A. Proposed framework

It is challenging to construct effective and cost-efficient Intrusion Detection System (IDS) models due to the complexity of high traffic and the requirement to strike a balance between a high detection rate and inexpensive processing expenses. The result is a classifier that this research presents that is compatible with FSA. It is possible to decrease processing costs while enhancing intrusion detection system (IDS) detection rates thanks to its adaptable and functional design. The primary objective of the system is to minimize calculations while obtaining very precise answers. The five main processes of the proposed framework are as follows, as shown in Figure 5.1: dataset collection, data pre-processing, FSA, model construction and assessment, and analysis and selection. In what follows, we'll discuss each stage in more detail.

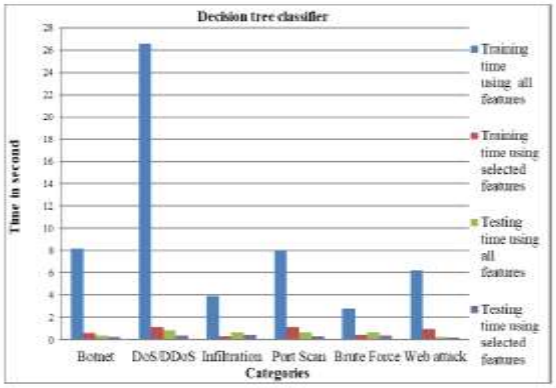


Figure 5.9 Training and testing times for DT classifier over CICIDS2017 utilizing both the RFE-selected features and all features across all categories

CONCLUSION

In order to develop an effective intrusion detection system, this chapter examines several classifiers that combine many FSTs. The study's findings reveal that reducing the amount of datasets in IDS achieves two objectives: first, it boosts the model's performance, and second, it achieves another target. handling-related expenses are decreased. A DT classifier that employs RFE as FST outperforms its FSA counterparts on the NSL-KDD dataset. The U2R assault group is the only exception in this regard. I couldn't agree with you more on the F-measure, memory, accuracy, and precision. Its operation is different from that of other algorithms that use FSA. Additionally, a more refined and compact set of traits has been discovered using the proposed FST. Methods for ranking and data gain for the models allowed us to achieve this. The provided FST was used for this purpose. The study's findings indicate that the NSL-KDD dataset has thirteen crucial features, whereas the CICIDS 2017 dataset contains eight vital features. By minimizing the number of features used by the model, we might potentially enhance its performance while decreasing the computing resources needed. Evaluations were conducted to analyze the recall, G-means, precision, sensitivity, F-measure, accuracy, training time, and testing time of the RFE+DT model using the Realtime dataset (CICIDS2017). The model's superiority and efficacy were shown by comparing it to other well-known models that had previously been discussed. Researchers have proven that using Decision Trees (DT) for classification and Recursive Feature Elimination (RFE) for feature selection (FST) improves results while reducing computational load, according to many studies.

REFERENCES

[1]. Zhang, Y., Li, P., & Wang, X. (2019). Intrusion detection for IoT based on improved genetic algorithm and deep belief network. IEEE Access, 7, 31711-
[2]. Elmasry, W., Akbulut, A., & Zaim, A. H. (2020). Comparative evaluation of different classification techniques for masquerade attack detection. International Journal of Information and Computer Security, 13(2), 187-209.
[3]. Shelke, M. P. K., Sontakke, M. S., & Gawande, A. D. (2012). Intrusion detection system for cloud computing. International Journal of Scientific & Technology Research, 1(4), 67-71.
[4]. Rajput, D., & Thakkar, A. (2019). A survey on different network intrusiondetection systems and countermeasure. In Emerging Research in Computing Information, Communication and Applications: ERCICA 2018, Volume 2 (pp497-506). Springer Singapore.
[5]. Wang, C., Zhao, T., & Liu, Z. (2020). An activity theory model for dynamic evolution of attack graph based on improved least square genetic algorithm. International Journal of Information and Computer Security, 12(4), 397-415.
[6]. Larson, D. (2016). Distributed denial of service attacks–holding back the flood. Network Security, 2016(3), 5-7.
[7]. Vijayakumar, D. S., & Ganapathy, S. (2022). Multistage ensembled classifier for wireless intrusion detection system. Wireless Personal Communications, 122(1), 645-668.
[8]. Alkasassbeh, M. (2017). An empirical evaluation for the intrusion detection features based on machine learning and feature selection methods. arXiv preprint arXiv:1712.09623.
[9]. Gu, S., Cheng, R., & Jin, Y. (2018). Feature selection for high-dimensional classification using a competitive swarm optimizer. Soft Computing, 22, 811- 822.
[10]. Rao, H., Shi, X., Rodrigue, A. K., Feng, J., Xia, Y., Elhoseny, M., ... & Gu, L. (2019). Feature selection based on artificial bee colony and gradient boosting decision tree. Applied Soft Computing, 74, 634-642.

[11]. Mafarja, M., Aljarah, I., Faris, H., Hammouri, A. I., Ala'M, A. Z., & Mirjalili, S.(2019). Binary grasshopper optimisation algorithm approaches for feature selection problems. Expert Systems with Applications, 117, 267-286.
[12]. Thanh, H., & Lang, T. (2019). An approach to reduce data dimension in building effective network intrusion detection systems. EAI Endorsed Transactions on Context-aware Systems and Applications, 6(18).
[13]. Almseidin, M., Alzubi, M., Kovacs, S., & Alkasassbeh, M. (2017, September).Evaluation of machine learning algorithms for intrusion detection system.In 2017 IEEE 15th International Symposium on Intelligent Systems andInformatics (SISY) (pp. 000277-000282). IEEE.
[14]. Kok, S. H., & Abdullah, A. NZJhanjhi, and Mahadevan Supramaniam. A review of intrusion detection system using machine learning approach. International Journal of Engineering Research and Technology, ISBN 0974, 3154(12), 1.
[15]. Al-Jarrah, O. Y., Siddiqui, A., Elsalamouny, M., Yoo, P. D., Muhaidat, S., & Kim, K. (2014, June). Machine-learning-based feature selection techniques for large-scale network intrusion detection. In 2014 IEEE 34th international conference on distributed computing systems workshops (ICDCSW) (pp. 177- 181). IEEE.