International IT Journal of Research (IITJR), ISSN: 3007-6706
Volume 3, Issue 2, April-June, 2025
Available online at: https://itjournal.org/index.php

An AI-Driven Intrusion Detection System Was Put Into Place to Safeguard Cloud Infrastructure Against Potential Threats.

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

One of the most prominent examples of a contemporary intrusion detection system is the Intrusion Detection System (IDS). Preventing unwanted access to networks, safeguarding sensitive data, and creating an additional layer of security are key objectives of this program. Hosts and networks are safeguarded from danger by intrusion detection systems (IDS), which check all network traffic for harmful material and notify administrators of any suspicious behavior. And alarm systems may be set to go off when they sense something out of the norm. A plethora of new markets have mushroomed thanks to the meteoric rise of the Internet. Examples of such developing industries include big data, cloud computing, and the internet of things (IoT). A probable cause of the increase in attack frequency might be the network-wide surge in data production and transmission speeds. This is why a large number of researchers have focused on improving intrusion detection systems (IDS) to protect against attacks and other threats associated with them. There is a good chance that most of the data contained in network logs includes attributes that are irrelevant to the identification or classification of attacks. Therefore, professionals still have a hard time making sense of this kind of network data and figuring out if the chosen characteristics could make IDSs more effective. In addition, a big collection is necessary for breach detection systems to manage the diverse range of threats. To improve the accuracy and speed of the intrusion detection system (IDS), it is necessary to determine the main characteristics, which is a difficult but essential stage.

The current intrusion detection system (IDS) uses a variety of deep learning, evolutionary, and machine learning algorithms to spot threats and gain insight from past data by analyzing patterns.

These solutions are likely to be costly to implement since they take into account every aspect of traffic at the same time. Still, the outcomes produced by these methods are commendable. Therefore, it's an ongoing challenge to find ways to save expenses without compromising efficiency or providing features that aren't necessary. However, in an effort to address these issues, FSA analysis was first carried out on the NSL-KDD and CICIDS2017 files. We did this to eliminate superfluous qualities and zero down on the most crucial ones. This need necessitated the development of more affordable intrusion detection systems (IDS) capable of operating in very large networks. We examine and evaluate many models that use FSA with NSL-KDD datasets to enhance the intrusion detection system's (IDS) detection engine.

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

INTRODUCTION

Over the last several decades, internet access has become ubiquitous in today's world. The majority of our customers use a variety of electronic devices, including smartphones, computers, tablets, and more, to access our services around the clock and from anywhere in the world. Because of this, crucial or sensitive information could be sent across these networks. Another consequence of the ever-evolving internet is the constant movement of private data between devices and data centers for the purposes of archiving and retrieval.

These outcomes provide an opening for the assailants to conduct many assaults, each of which poses a threat to the designated target. An attacker might potentially take advantage of system security flaws using a range of cutting-edge approaches. If unauthorized individuals get access to the system, it might jeopardize it and cause sensitive information to be disclosed or their accounts to be breached. System administrators and security staff must protect themselves from modern threats using modern security solutions. Big data and the Internet of Things are two examples of the new technologies that are adding to the deluge of data.

As a consequence, the network experiences increased data congestion, which in turn makes changing the attack profile slower, more complex, and more problematic. The capacity to extract useful information from large datasets is an additional crucial talent for data scientists, businesses, and marketers. The massive volume of data produced by these links is beginning to get the attention of academics and scientists concerned about network security. The main cause of this is the ever-growing number of people using the internet. Network security is the study of identifying and resolving security vulnerabilities in order to prevent unauthorized individuals from obtaining access to computer systems or networks. Attacks such as denial-of-service (DoS), user-to-root (U2R), remote-to-local (R2L), probing, and countless more have prompted the development of several defensive solutions throughout the last two decades. There are several instances, such as firewalls and antivirus programs. In order to identify new types of attacks and harmful data or traffic that might harm the system or network, basic security measures must be established. Such a device is known as an intrusion detection system (IDS) [1]. The common term for them is "IDSs." An intrusion detection system (IDS) uses a mix of hardware and software to collect, evaluate, and identify incoming data. These technologies may help find and remove hazards including fraudulent attacks, possible dangers, and bothersome systems on both the network and in individual computers [2]. One of the functions of an intrusion detection system (IDS) is to safeguard confidential information as it travels across a network. Examining the intrusion detection system (IDS), analyzing its data using mathematical or statistical techniques, and ensuring that it warns network administrators and managers of any suspicious behavior are all necessary to perform these duties and meet these goals [3].

A. AN OVERVIEW OF THE PROBLEM

Having access to these services has become more important in the last 20 years due to the emergence of COVID-19 and other internet-based threats. The rapid rise to prominence could be attributed to the introduction of much improved Internet technologies. Computers, tablets, smartphones, and other such electronic devices allow users to access these services swiftly and from any location. This implies that networks are seeing an increase in the transfer of sensitive data between computers and data storage facilities. Because of this, criminals may easily conduct large-scale attacks, putting the firm or its consumers at risk as they try to circumvent security safeguards. Cybercriminals use a myriad of sophisticated strategies in their quest to breach computer systems' defenses. The article goes into depth about a few of these methods. Theft of user accounts, exploitation of sensitive data, or illegal access to the system could result from this. Data security and network strengthening are two areas where experts and academics are concentrating their efforts to mitigate the effects of these assaults. A solution has been given by the widespread use of intrusion detection systems, or IDS. As information is being submitted, intrusion detection systems check to see whether it concerns behaviors that affect the whole system or the entire network. Internet, social media, and the Internet of Things have all contributed to a dramatic growth in the amount of data generated and sent inside the network. The widespread use of these tools has led to this. Some of the possible side effects of network traffic can be rather annoying, while others might be completely negligible. To address this, effective intrusion detection systems (IDS) will include several monitoring approaches along with feature addition and removal tools. There is no way to overstate its importance in preventing the system's processing power and working time from increasing. This is why there are models for reducing or removing features and rapid decision-making engines [17]. Using a single classifier or estimate to evaluate and compare many models is probably not the best strategy.

2. BACKGROUND

It is becoming more difficult than ever to maintain computer system security due to the ever-changing nature of attack strategies and the growth of new kinds of networks (such as wireless sensor networks and software-defined). Upon initial establishment of these more recent networks, security precautions were not given primary importance. Traditional security protocols often fail to provide sufficient protection for such networks. Given this reality, it is critical to have a fast security system that can identify attempts to access a computer system. In response to a need, what is now generally recognized as an intrusion detection system (IDS) was developed. infiltration detection systems (IDS) mainly monitor networks for any signs of infiltration [19]. This is done to ensure that the three pillars of computer security—authenticity, integrity, and confidentiality—remain undamaged. The goal is to identify any possible violations or threats.

Obtaining Knowledge Through: The first thing that needs to be done in order to implement this strategy is to pinpoint the specific system that will be the target of the attack. To do this, you may make use of search tools or follow the directions provided by the network. Using network commands such as "nslookup" to obtain information about hosts and servers, including the domain name, it is feasible to intercept and steal data that is being sent over the internet. Section III: Details of the Packet Analysis Summary: To phrase it another way, this strategy may include "checking the network packet for the improper use of sensitive information." An R2L hack might take place in the event that an unauthorized person is able to obtain access to the system. Obtaining access may be accomplished via a variety of means, including the scanning of data files in order to get login credentials or the remote control of the system through the use of malicious software (often a trojan horse). In most cases, these vulnerabilities can only be exploited when the machine that is

being targeted has a limited number of open ports. Following the completion of the analysis of the packet data, the fourth step entails the identification and recording of the particular patterns and signatures that are connected with a wide range of known security vulnerabilities that might be exploited by individuals present inside the organization. The fingerprints and patterns may be preserved in order to attain this goal. Due to the fact that these signs and patterns are saved in the database for future reference, the security guard is able to promptly report any suspicious behavior.

5. Sending a sign of caution: Immediately after the attack pattern is identified, an alert is sent to the management of the security system. It is necessary to send out a warning signal if a signature or pattern is found to be identical.

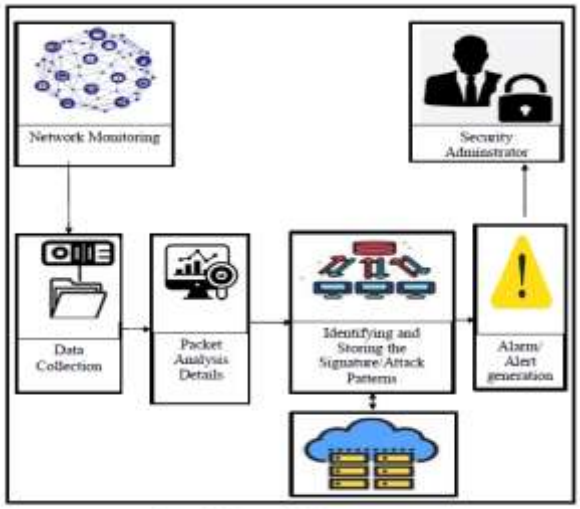

Figure 2.3 Advanced IDS's component.

A. CATEGORIZATION OF IDENTIFIERS

The intrusion detection system (IDS) sets the locations and information sources it monitors after giving the topic substantial consideration. Finding abnormalities, implementing detection processes, and considering security considerations led to the development of several intrusion detection systems.

Infrastructure, instruments, or systems for responding might be involved.

Understanding the deployment, operation, and purpose of intrusion detection systems (IDS) is critical for identifying and avoiding security threats, hence it's important to categorize them. Intruder detection systems keep a close eye on networks and systems in order to spot any signs of suspicious activity, policy violations, or malicious intent. Experts in the field of security can classify everything into relevant categories, allowing them to address any problem. An intrusion detection system's (IDS) architecture, deployment, threat detection capabilities, data analysis time, and response strategy might provide the foundation for grouping these systems into different types. Each grouping highlights a unique aspect of the intrusion detection system (IDS), which together provide more thorough security monitoring.

One of the most important parts of IDS categorization is finding the deployment place with great precision. Security of files, user logins, configuration changes, application logs, system calls, and host-based intrusion detection systems are all monitored by separate hosts or servers. An advanced intrusion detection system (HIDS) can identify hidden dangers, illegal access, and harmful actions that do not compromise data that is available to the public on a network. Manage HIDS across several locations, however, could be challenging and resource-intensive. However, by positioning themselves strategically inside a network, Network-Based Intrusion Detection Systems (NIDS) can keep tabs on device-to-device communication. In order to detect potential dangers to networks, network intrusion detection systems (NIDS) analyze packet contents and headers. Port scanning, virus transmission, and denial-of-service attacks are all examples of such dangers. Network intrusion detection systems (NIDS) may safeguard several systems simultaneously, but they're not foolproof. Encrypting crucial data or targeting only one server might make them ineffective. To increase the sensitivity and accuracy of its detections, hybrid intrusion detection systems (IDS) combine host-based and network-based approaches.

B. IDS'STAXONOMY

Intrusion detection systems (IDS) monitor networks and other systems for suspicious activity and alert administrators immediately, as the name implies. A careful eye on the system allows us to do this. In their work, Liao et al. [31] categorize

intrusion detection systems according to four main features. Considerations include information availability, system instability, detecting mechanism, and reaction time. Referring to reference [31], Figure 2.4 shows the various Intrusion Detection Systems (IDSs) that are mentioned.

To better understand and work with intrusion detection systems (IDSs), we may categorize them according to their use, attack detection capabilities, data source, response mechanisms, and system architecture. By perusing this section, experts, designers, and security managers may find the intrusion detection system (IDS) options most suited to their organization's demands, network capacity, threat profiles, and performance standards. In the same way that cyber threats have evolved in complexity, so too has the terminology used to describe intrusion detection systems. This leads to the evolution of detection systems that are multi-modal and increasingly intricate.

Categorizing intrusion detection systems (IDSs) according to their deployment site is a crucial feature of IDS classification since it indicates the system component of which the IDS is a part. Every host or server location has its own host-based intrusion detection system (HIDS) that watches system calls, application logs, file system changes, and user activity. By using HIDS, it is feasible to simultaneously detect attempts at insider attacks, higher privileges, and undesired file modifications.

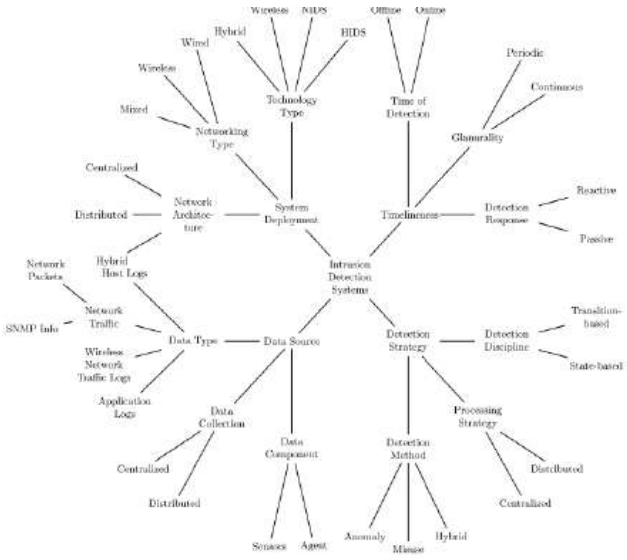

Figure 2.4. IDS's taxonomy suggested by [31].

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.

LITERATURESURVEY

Data scientists and information researchers still haven't found a solution to the big mystery of dimensionality reduction. In order to facilitate dimensionality reduction in very large datasets, a multitude of IDS models have been created within the last few decades. Every network, including KDD99 and NSL KDD, is available here. Various studies have used the FSA to enhance intrusion detection systems (IDS) and sidestep issues related to high-dimensional data. Data simplification and task management remain obstacles for professionals, albeit [43]. With the exponential expansion of network traffic, the number

of possible threats has also expanded. Because of this, a number of experts have implemented several Intrusion Detection System (IDS) machine learning strategies based on FSAs.

Mukkamala and colleagues [44] used Support Vector Machines (SVM) and neural networks (NN) to evaluate intrusion detection systems. Support Vector Machines (SVMs) shown exceptional efficiency and adaptability when confronted with large datasets throughout the tests. A significant time commitment from NN is required for learning. Fleuret et al. (2004) used a technique known as the joint information approach to ascertain which qualities were pertinent to this conversation. Combining SVM with a Bayes network improves performance over SVM alone. The majority of their studies have focused on total work hours [45]. Chebrolu et al. (2005) investigated intrusion detection systems (IDS). The investigation made use of cutting-edge technological developments such as reverse classification trees, Bayes networks, and others. Their strategy yielded twelve crucial qualities that they used to effectively identify and escape several types of attacks. There has been evidence of surprisingly high detection rates of U2R attacks [46]. Chou et al. (2008) used many new feature selection algorithms (FSAs), such as rapid CFS and correlation-based feature selection (CFS), to solve problems with multi-dimensional data. The data are repetitive and lack specificity, which is a problem.

4. A FORMAL APPROACH TO EMBEDDED LEARNING-BASED INTRUSION DETECTION SYSTEMS

In order to set the stage for FST-based systems, this chapter identifies critical components that might enhance the performance of the recognition engine. Numerous algorithms have enhanced their performance by using recursive feature elimination (RFE).

Attainment of the objective has been accomplished with all required characteristics accompanied. The NSL KDD Dataset is used for both the development and assessment of our methods. In this way, you can show how picking the correct features improves accuracy compared to picking any features at all. In this study, we evaluate RFE in comparison to various ensemble classifiers, including RF, GB, AB, and ET. Extra classifiers are also considered. The main function of these algorithms is to sort the data into several groups. From what we can see from our comparisons, picking the right features is the secret to significantly improving the classifier's success rate and overall performance. A summary of the main sources I used for this chapter is provided here.

After running the UFS, the components' significance was determined using the RFE technique, the ANOVA F-Test, and the select_Percentile tool.

A. COLLECTION DATA

After its creation at MIT's Lincoln Labs in 1999, the KDD 1999 dataset has seen heavy use by academics over the last 20 years [89]. Improved upon already by the improvements made to the NSL-KDD dataset, the KDD1999 dataset is now much better (81). It is anticipated that this collection will address several issues.

The following is one way to describe the KDD 1999 dataset:

The objective results produced by our algorithms, which draw from a wide range of approaches, may lead to improved identification rates or accuracy when applied to regular data. This may be feasible since there is no duplicate data in either the training set or the testing set. The original KDD 1999 dataset is comprised of two distinct components: the training dataset and the testing dataset. Each component contains a total of records. Doing so removes the potential of any given piece of data being duplicated.

Use of the NSL-KDD dataset is strongly recommended for several reasons, such as: In order for the algorithm to provide more impartial results, it is critical to exclude comparable data. The testing dataset and the training dataset both include a large number of events. Instead of randomly selecting pieces from an ever-decreasing set, it's feasible to test the whole collection. Finally, it has a ton of features, such as precise network design, comprehensive packet capture, structured notes, and a ton of other advantages.

International IT Journal of Research (IITJR), ISSN: 3007-6706
Volume 3, Issue 2, April-June, 2025
Available online at: https://itjournal.org/index.php

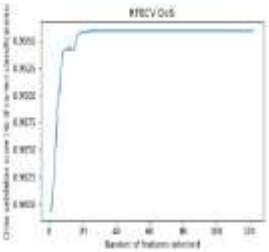

Figure 4.3 DoS RFECV with AB

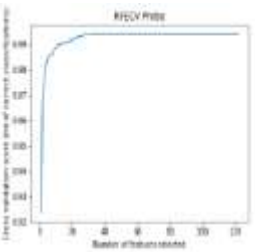

Figure 4.4 PROBE RFECV with AB

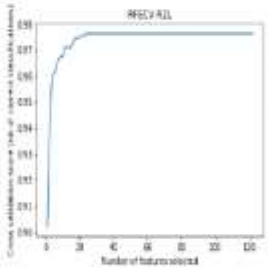

Figure 4.5 R2L RFECV with AB

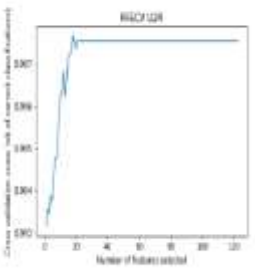

Figure 4.6 U2R RFECV with AB

5. ARCHITECTURE FOR ADAPTIVE IDS.

Discover a fresh approach to using FSA in this chapter that will assist you in identifying important features and eliminating unnecessary ones. In addition, it confirms that the NSL-KDD datasets are really IDS sensitive.

The engine went through a series of experiments to find the most accurate predictor. By using the CICIDS2017 real-time dataset, we were able to assess the top FST and predictor. Testing examples are provided in this chapter to demonstrate how the proposed model's core features enhance IDS performance while significantly decreasing processing needs. Applying the proposed model improved accuracy by 99.21% on the NSL-KDD dataset and by 99.94% on the CICIDS2017 dataset. In order to accomplish both of these goals, testing was necessary..

A. PRESENTED STRUCTURE

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.

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.