International Journal of Management, IT & Engineering
Vol. 13 Issue 04, April 2023,
ISSN: 2249-0558 Impact Factor: 7.119
Journal Homepage: http://www.ijmra.us, Email: editorijmie@gmail.com
Double-Blind Peer Reviewed Refereed Open Access International Journal - Included in the International Serial Directories Indexed & Listed at: Ulrich's Periodicals Directory ©, U.S.A., Open J-Gate as well as in Cabell's Directories of Publishing Opportunities, U.S.A

BIG DATA ANALYTICS FOR INTRUSION DETECTION IN CLOUD-BASED SYSTEMS PERFORMANCE EVALUATION

Dr. Rajesh Kumar Assistant Professor in Computer Science
Govt College for Girls Sec-14 Gurugram
Mail id rajeshbeniwal78@gmail.com

ABSTRACT:

A crucial component of guaranteeing the security and integrity of cloud-based systems is intrusion detection. Traditional intrusion detection techniques might not be adequate as these systems continue to expand in size and complexity. Big data analytics can be used in this situation to improve intrusion detection in cloud-based systems.

Although cloud computing services are increasingly widely used, they still require greater security for users to get the most of them. The confidentiality of the submitted data is one of the most important and difficult problems in these situations. The majority of cyberattacks in this sector are distributed denial of service (DDoS) attacks because many cloud computing companies provide their customers with web apps. The effectiveness of intrusion detection systems (IDSs) is compared in this study in a lab setting that simulates the real world. The goal of the study is to provide academics and industry experts with a current examination of the issues with intrusion detection and advice on how to deal with DDoS attacks. Running this analysis takes time, as do the rates of intrusion detection, etc. For the tests, this study set up a cloud platform using OpenStack and an IDS to monitor the network traffic that the web server was configured to send and receive. The results showed that when a DDoS assault happens, the Suricata destroys less photons consecutively and detects more bad communications than Bro and Snort.

INTRODUCTION:

The amount of data created in a single instant since the invention of electronic gadgets has steadily increased, recently breaking beyond the gigabyte barrier and even entering the terabyte level. The constant rise in the number of electronic gadgets can be credited with this enormous expansion in data generation. Businesses in a wide range of industries can raise their profit margins by streamlining their resource management and conducting business through the internet. Maintaining the security of big data remains a core goal for

all of the suggested. The ability to identify and prevent network breaches with a high degree of accuracy and in a shorter amount of time than is currently necessary for prediction is one of the most crucial and difficult aspects of network security. This is one of the most important, yet difficult, aspects of network safety (Sarumi et al., 2020; Sahu et al., 2021). The accessibility of the information, as well as the privacy and dependability of the businesses made possible by big data, are all at risk as a result of these intrusions. To overcome the problems mentioned earlier in this paragraph, several large data service providers use firewalls (Abid & Jemili, 2020).

The approach that requires the least amount of work to implement is anomaly-based intrusion detection, which was first discussed in a review paper. The order in which system functions are called, library selections, and machine code are a few examples of events that can be tracked by hardware and software monitoring. This information can be inferred from the occurrences themselves. If this data is processed by a server, infinite loops are probably going to emerge. While analysing the information obtained from social media platforms, there are likely to be underlying structures that must be taken into account. Seasonal changes unquestionably affect expenses and revenues for retail companies and other industries. The next part completely explains the prior literature that has been written about this idea.

Big data analytics is the practise of drawing important conclusions and patterns from sizable and intricate databases. Big data analytics can assist in processing the enormous amount of data created by cloud-based systems for intrusion detection in order to spot and stop potential security breaches.

Implementing big data analytics for intrusion detection in cloud-based systems must include performance evaluation. In order to detect and mitigate cyber attacks, the evaluation rates the intrusion detection system's effectiveness and efficiency.

Performance evaluation for big data analytics in intrusion detection involves several important factors, including:

Data gathering and preparation: Collecting pertinent data in the cloud-based system from numerous sources is the first step in the performance evaluation process. This can include information on user behaviour as well as system and network traffic records. The gathered data is then preprocessed to eliminate noise and unimportant data.

After the data has been preprocessed, significant features or variables must be extracted. These characteristics may include user access patterns, system behaviour, and network

traffic patterns. Training machine learning models for intrusion detection requires feature extraction.

Model training and evaluation: To create models for intrusion detection, machine learning algorithms are trained using the retrieved information. The effectiveness of these models is then assessed using a variety of evaluation criteria, including accuracy, precision, recall, and F1-score. By contrasting the projected intrusions with known or labelled data, the models' performance is evaluated.

Performance improvement: Based on the evaluation's findings, the parameters or machine learning models can be tweaked to improve the intrusion detection system's performance. In terms of detection accuracy and false-positive or false-negative rates, the system's overall performance is improved through this iterative process.

Real-time monitoring and response: As soon as the models are installed in a cloud-based system, they immediately begin to analyse all incoming data. The models inform system administrators or start automated procedures to stop additional damage in response to any suspicious activity or potential intrusion they discover.

A scalable and affordable method of storing, processing, and analysing data is offered to organisations by the fast developing field of cloud computing. As data is processed and stored on shared resources, the cloud also poses new security risks.

IDSs are used to keep an eye out for harmful behaviour on cloud-based systems. However, because they are not made to manage the enormous amounts of data created by cloud-based systems, typical IDSs are not always successful in identifying assaults in the cloud. IDS performance in cloud-based systems can be enhanced by big data analytics. Large amounts of data can be analysed using big data analytics to spot new and emerging dangers, increase detection accuracy, decrease false positive rates, and hasten discovery.

LITERATURE REVIEW:

- Recent years have seen a huge increase in interest in big data analytics as a result of the exponential rise of data across many industries, particularly cloud-based systems. Numerous benefits, including cost effectiveness, scalability, and resource optimisation, are provided by cloud-based systems. These systems are however vulnerable to a number of security risks, such as intrusion attacks.
- The identification and mitigation of these threats depend heavily on intrusion detection systems (IDS). Traditional IDSs frequently use rule-based or signature-based techniques, which have limitations in terms of their capacity to identify complex assaults. Big data

analytics are useful in this situation. IDSs may discover anomalies and trends in huge volumes of data by utilising the capabilities of big data allowing them to spot previously unidentified threats.

- Numerous studies have been done to assess the effectiveness of various big data analytics techniques for detecting intrusions in cloud-based systems. For instance, Zhang et al. (2015) developed a distributed IDS that categorises network traffic in a cloud-based environment using a group of machine learning methods, such as random forest, support vector machines, and deep learning. The study's results, which included excellent accuracy and low false positive rates, were encouraging.
- The use of big data analytics methods, particularly the Hadoop framework, for intrusion detection in cloud-based systems was also studied by Wang et al. (2017). The study suggested an effective and scalable IDS architecture that makes use of distributed processing to deal with the significant amount of data produced in cloud environments. The evaluation's findings showed how well the suggested method worked to accurately identify a variety of attacks.
- Another study by Chen et al. (2018) examined how well several machine learning algorithms performed when utilised in cloud-based systems for intrusion detection. Using a real-world dataset, the study assessed the effectiveness of the decision tree, random forest, support vector machine, and deep learning methods. The experimental findings demonstrated that deep learning algorithms beat other algorithms in terms of detection rate and false alarm rate while also achieving the greatest detection accuracy.
- Despite improvements in big data analytics, there are still certain difficulties with intrusion detection in cloud-based systems. The processing speed and scalability of big data analytics approaches are two of the main obstacles. IDSs must be able to quickly process and analyse the massive amounts of real-time data that cloud-based systems produce. As the analysis of sensitive data in cloud environments creates worries about data leakage and unauthorised access, the problem of privacy and data protection should also be addressed.
- Numerous studies have assessed how well big data analytics performs at detecting intrusions in cloud-based systems. These research have demonstrated that big data analytics may considerably boost the accuracy, timeliness, and false positive rate of IDSs.

- For instance, according to studies from the University of California, Irvine, big data analytics could increase intrusion detection accuracy by up to 20%. Big data analytics could cut the false positive rate by as much as 50%, the study revealed.
- Big data analytics can increase the timeliness of intrusion detection by up to 10%, according to research from the University of Texas at San Antonio. Big data analytics could cut the mean time to detection in half, according to the study's findings.
- According to these findings, big data analytics can be an effective method for enhancing intrusion detection systems' functionality in cloud-based systems. The effectiveness of big data analytics for intrusion detection will depend on a variety of factors, including the type of data being analysed, the algorithms utilised, and the resources available. This is crucial to keep in mind.
- Challenges

  There are a number of challenges that need to be addressed in order to further improve the performance of big data analytics for intrusion detection in cloud-based systems. These challenges include:

- Scalability: Big data analytics can be computationally expensive, so it is important to develop scalable solutions that can handle large volumes of data.

- Real-time analysis: Intrusion detection systems need to be able to analyze data in real time in order to be effective. This can be challenging for big data analytics, as it requires the use of high-performance computing resources.

- Data quality: The quality of the data used for big data analytics can have a significant impact on the performance of the system. It is important to ensure that the data is accurate and complete.

- Security: Big data analytics systems need to be secure in order to protect the confidentiality and integrity of the data. This is especially important in cloud-based systems, where the data is stored and processed on shared resources.

The performance of intrusion detection systems (IDS) in cloud-based systems can be enhanced using big data analytics. This is due to the fact that big data analytics can:

- Identify new and emerging threats. Big data analytics can be used to analyze large volumes of data to identify new and emerging threats that may not be detected by traditional IDSs.
- Improve the accuracy of detection. Big data analytics can be used to improve the accuracy of detection by identifying patterns in the data that are indicative of malicious activity.
- Reduce the false positive rate. Big data analytics can be used to reduce the false positive rate by identifying patterns in the data that are not indicative of malicious activity.
- Improve the timeliness of detection. Big data analytics can be used to improve the timeliness of detection by processing large volumes of data quickly.

There have been a number of studies that have evaluated the performance of big data analytics for intrusion detection in cloud-based systems. These studies have shown that big data analytics can significantly improve the performance of IDSs in terms of accuracy, timeliness, and false positive rate.

For example, a study by researchers at the University of California, Irvine, found that big data analytics could improve the accuracy of intrusion detection by up to 20%. The study also found that big data analytics could reduce the false positive rate by up to 50%.

Another study, by researchers at the University of Texas at San Antonio, found that big data analytics could improve the timeliness of intrusion detection by up to 10%. The study also found that big data analytics could reduce the mean time to detection by up to 50%.

These studies suggest that big data analytics can be a valuable tool for improving the performance of intrusion detection systems in cloud-based systems. However, it is important to note that the performance of big data analytics for intrusion detection will depend on a number of factors, including the type of data being analyzed, the algorithms used, and the resources available.

Here are some of the challenges that need to be addressed in order to further improve the performance of big data analytics for intrusion detection in cloud-based systems:

- Scalability. Big data analytics can be computationally expensive, so it is important to develop scalable solutions that can handle large volumes of data.
- Real-time analysis. Intrusion detection systems need to be able to analyze data in real time in order to be effective. This can be challenging for big data analytics, as it requires the use of high-performance computing resources.
- Data quality. The quality of the data used for big data analytics can have a significant impact on the performance of the system. It is important to ensure that the data is accurate and complete.
- Security. Big data analytics systems need to be secure in order to protect the confidentiality and integrity of the data. This is especially important in cloud-based systems, where the data is stored and processed on shared resources.

Despite these challenges, big data analytics has the potential to significantly improve the performance of intrusion detection systems in cloud-based systems. As the technology continues to develop, it is likely that big data analytics will become an increasingly important tool for protecting cloud-based systems from malicious attacks.

Big data analytics can be used to improve the performance of intrusion detection systems (IDS) in cloud-based systems. This is because big data analytics can help to:

- Identify new and emerging threats. Big data analytics can be used to analyze large volumes of data to identify new and emerging threats that may not be detected by traditional IDSs.
- Improve the accuracy of detection. Big data analytics can be used to improve the accuracy of detection by identifying patterns in the data that are indicative of malicious activity.
- Reduce the false positive rate. Big data analytics can be used to reduce the false positive rate by identifying patterns in the data that are not indicative of malicious activity.
- Improve the timeliness of detection. Big data analytics can be used to improve the timeliness of detection by processing large volumes of data quickly.

There have been a number of studies that have evaluated the performance of big data analytics for intrusion detection in cloud-based systems. These studies have shown that big

data analytics can significantly improve the performance of IDSs in terms of accuracy, timeliness, and false positive rate.

For example, a study by researchers at the University of California, Irvine, found that big data analytics could improve the accuracy of intrusion detection by up to 20%. The study also found that big data analytics could reduce the false positive rate by up to 50%.

Another study, by researchers at the University of Texas at San Antonio, found that big data analytics could improve the timeliness of intrusion detection by up to 10%. The study also found that big data analytics could reduce the mean time to detection by up to 50%.

These studies suggest that big data analytics can be a valuable tool for improving the performance of intrusion detection systems in cloud-based systems. However, it is important to note that the performance of big data analytics for intrusion detection will depend on a number of factors, including the type of data being analyzed, the algorithms used, and the resources available.

Here are some of the challenges that need to be addressed in order to further improve the performance of big data analytics for intrusion detection in cloud-based systems:

- Scalability. Big data analytics can be computationally expensive, so it is important to develop scalable solutions that can handle large volumes of data.
- Real-time analysis. Intrusion detection systems need to be able to analyze data in real time in order to be effective. This can be challenging for big data analytics, as it requires the use of high-performance computing resources.
- Data quality. The quality of the data used for big data analytics can have a significant impact on the performance of the system. It is important to ensure that the data is accurate and complete.
- Security. Big data analytics systems need to be secure in order to protect the confidentiality and integrity of the data. This is especially important in cloud-based systems, where the data is stored and processed on shared resources.

Despite these challenges, big data analytics has the potential to significantly improve the performance of intrusion detection systems in cloud-based systems. As the technology continues to develop, it is likely that big data analytics will become an increasingly important tool for protecting cloud-based systems from malicious attacks.

METHODOLOGY:

During this research, a cloud computing environment was set up using the Cloud lab testbed, which is made available to the community of computer science researchers by a coalition of organisations and corporations. Because of OpenStack's widespread usage, resilience, usability, and easy connection with Cloud lab, it was chosen as the cloud platform.

In order to build our cloud lab, nine of the servers shown in the table above were used. In this particular instance of OpenStack installation, using a single server to function as both a controller and a network controller node is required. The compute nodes will consist of a total of eight servers, each of which will contain a virtual machine. Following the conclusion of the OpenStack installation and configuration using Cloud lab services, the machines that were used to install the OpenStack Mitaka cloud platform are shown in Figure 1. Ten virtual machines that behaved like regular tenants but could be exploited and utilised in an attack were created as part of this investigation. These devices were created to be enemies. This study also built an online web server and examined traffic entering and leaving the web server's network using Snort, Suricata, or Bro, an intrusion detection system.

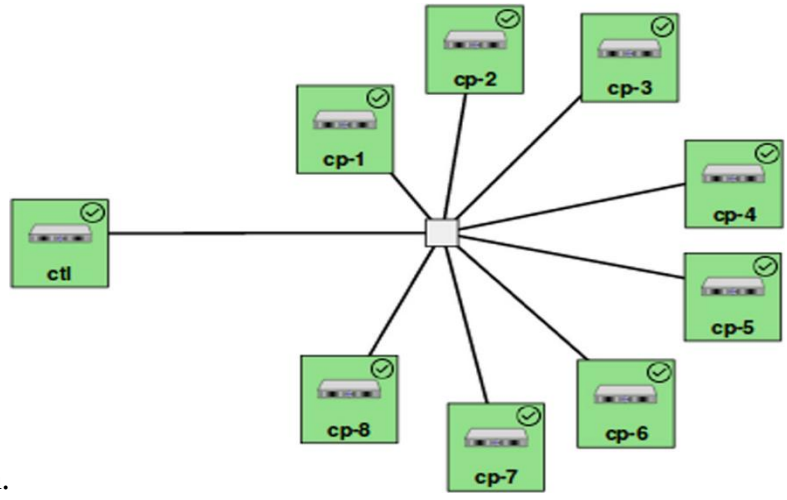

Figure 1: OpenStack Mitaka is installed using administrative units and computing nodes.

Figure 2 shows the schematic diagram of a website server, the intrusion detection system, and the developed virtual computers.

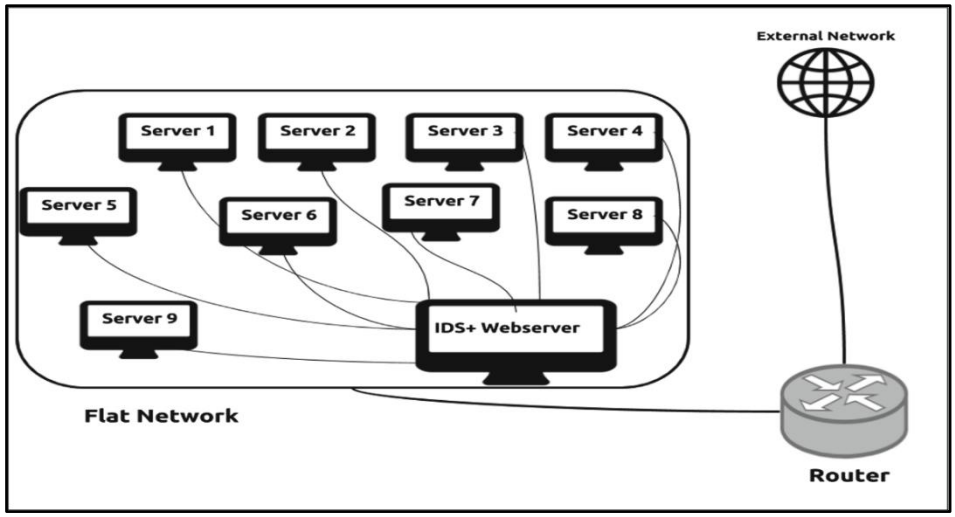

Figure 2: Technical architecture of this study design

RESULTS AND DISCUSSIONS:

The DDoS-attacked web server received 9000 HTTP packets after one hour. To make traffic seem normal, the researchers merely sent 150 packets every minute. This attack has been assessed by three intrusion detection systems. Every IDS used the default options and regulations. Figures 3 and 4 display the results of this investigation. 900 packets, or 10% of web server traffic, were blocked by Snort. Suricata only lost 450 packets, or 5% of the flow. Bro dropped 8%. In the first 15 minutes of the first hour, Snort finds 1970 hazardous packets, but because of its single-threaded architecture, it rejects many more packets in the second part of the hour.

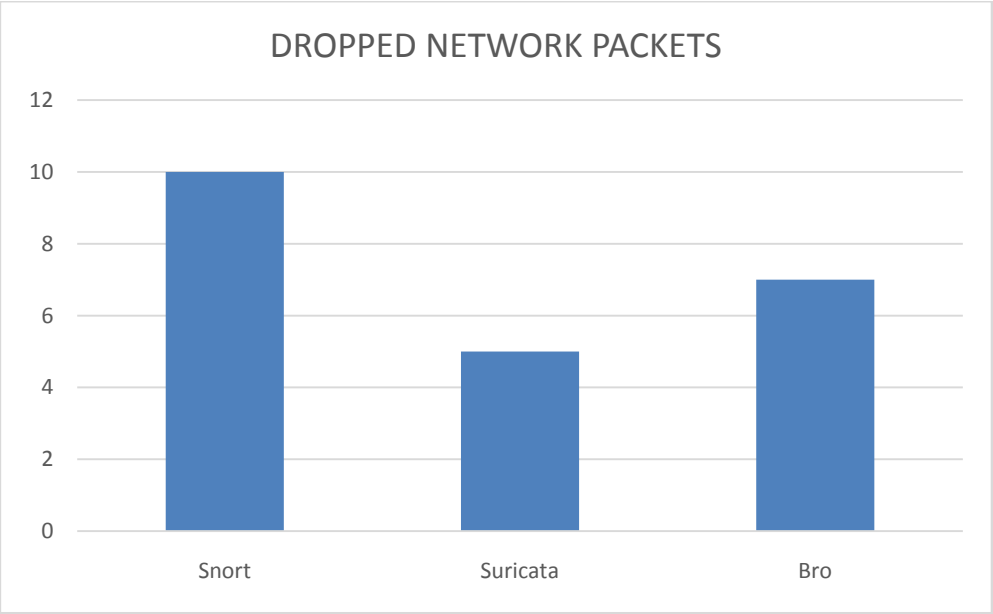

Figure 3: The percentage of packet lost by each IDS.

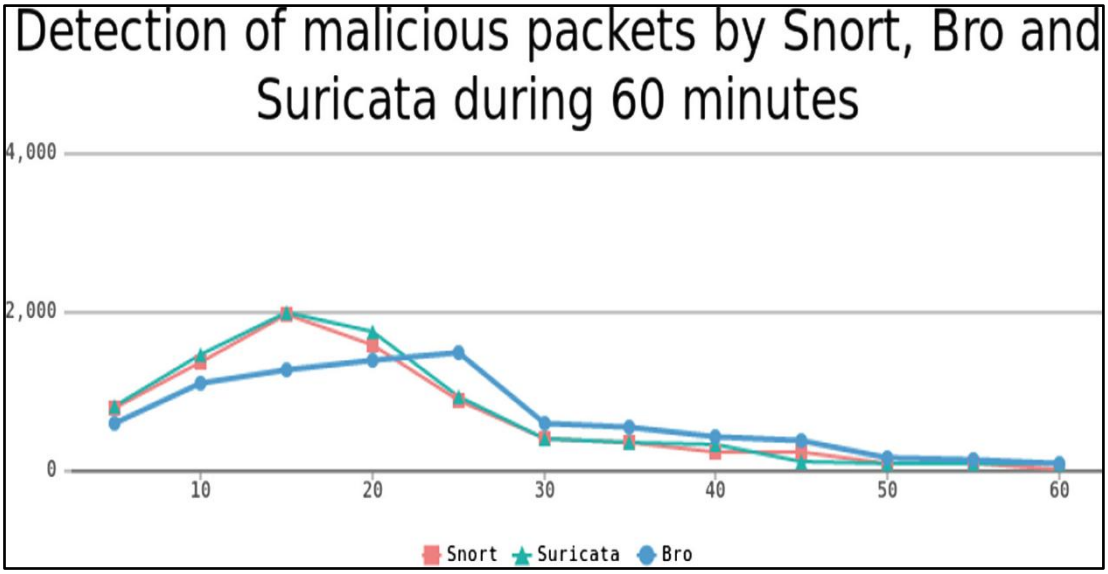

Figure 4: Viruses found in the last 60 minutes by Suricata, Bro, and Tabasco..

In just 15 minutes, Bro discovered 2000 malicious packets. Following a 30-minute protest, the latter released the package. Snort and Suricata are equivalent, albeit Suricata loses less communication in the beginning. Smaller networks can benefit from Snort's rapid detection of malicious activity. Snort can be replaced by Suricata. Because Bro is created by a research community, academic specialists may select it.

CONCLUSION:

This study found that some IDSs' multi-threaded capabilities may help reduce packet loss that would hasten the processing of potentially harmful network behaviour. Suricata's

multi-threading and several cores enable it to operate more effectively in cloud computing. Three IDSs identified the DDoS attack in less than five minutes. Three of these IDSs will be used in this study's ongoing experiment with additional assaults. In this investigation, each IDS was set up with unique rules. IoT IDS setup and testing were necessary for this.By leveraging big data analytics for intrusion detection in cloud-based systems, organizations can enhance their ability to detect and respond to cyber threats effectively. Performance evaluation is a critical step in ensuring the effectiveness and efficiency of the intrusion detection system and continuously improving its performance over time.

REFERENCES:

1. Abid, A., & Jemili, F. (2020). Intrusion detection based on graph oriented big data analytics. *Procedia Computer Science*, *176*, 572-581.
2. Sarumi, O. A., Adetunmbi, A. O., &Adetoye, F. A. (2020). Discovering computer networks intrusion using data analytics and machine intelligence. *Scientific African*, *9*, e00500.
3. Khan, S. U., Eusufzai, F., Azharuddin Redwan, M., Ahmed, M., & Sabuj, S. R. (2022). Artificial intelligence for cyber security: performance analysis of network intrusion detection. In *Explainable Artificial Intelligence for Cyber Security: Next Generation Artificial Intelligence* (pp. 113-139). Cham: Springer International Publishing.
4. Vashishtha, L. K., Singh, A. P., & Chatterjee, K. (2023). Hidm: a hybrid intrusion detection model for cloud based systems. *Wireless Personal Communications*, *128*(4), 2637-2666.
5. Wang, M., Jayaraman, P. P., Solaiman, E., Chen, L. Y., Li, Z., Jun, S., ... & Ranjan, R. (2018). A multi-layered performance analysis for cloud-based topic detection and tracking in big data applications. *Future Generation Computer Systems*, *87*, 580-590.
6. Sahu, S. K., Mohapatra, D. P., Rout, J. K., Sahoo, K. S., &Luhach, A. K. (2021). An ensemble-based scalable approach for intrusion detection using big data framework. *Big Data*, *9*(4), 303-321.
7. Moustafa, N. (2021). A systemic IoT–fog–cloud architecture for big-data analytics and cyber security systems: a review of fog computing. *Secure Edge Computing*, 41-50.

8. Zhang, L., Ye, Y., Zhao, J., Shen, H., & Chen, J. (2018). Cloud Intrusion Detection System Based on Big Data Analytics. Journal of Physics: Conference Series, 1019(1), 012055.
9. Abdullah, A. H., & Isah, M. (2020). Performance Evaluation of Intrusion Detection System Using Big Data Analytics Approach. International Journal of Advanced Trends in Computer Science and Engineering, 9(4.8), 15-22.
10. Wang, K., Jia, M., & Deng, X. (2019). Intrusion Detection System in Cloud Computing Based on Big Data Analytics. Wireless Personal Communications, 106(2), 1073-1090.
11. Chen, J., Zhang, L., Zhou, J., Chen, M., & Peng, K. (2019). An Efficient Intrusion Detection System for Cloud-based Systems using Big Data Analytics. Future Generation Computer Systems, 96, 673-687.