International Journal of Research in Engineering and Applied Sciences(IJREAS)
Available online at http://euroasiapub.org/journals.php
Vol. 12 Issue 9, September- 2022
ISSN(O): 2249-3905, ISSN(P) : 2349-6525 | Impact Factor: 8.202 |
Thomson Reuters ID: L-5236-2015

BIG DATA ANALYTCS FOR INTRUTION DETECTION IN CLOUD BASED SYSTEMS: A PERFORMANCE ANALYSIS

Agha Urfi Mirza[1], Dr. Navin Kumar[2]

[1,2]Department of Computer Science, Capital University, Koderma (Jharkhand)

ABSTRACT:
Services for cloud computing are already widely used, but they still need to be more secure for customers to use them effectively. One of the biggest and most challenging issues in these contexts is the privacy of the uploaded data. Due to the fact that many companies that use cloud computing offer their clients web applications, distributed denial of service (DDoS) attacks make up nearly all of the cyberattacks in this industry. This study compares the performance of intrusion detection systems (IDSs) in a lab environment representative of the real world. The study's objective is to give academics and professionals a current analysis of the problems with intrusion detection and how to handle DDoS attacks. This analysis takes time running, detection rates for intrusions, etc. In the trials, this study set up a cloud platform utilizing OpenStack and an IDS to watch over the web server's configured network traffic. The findings demonstrated that when a DDoS attack occurs, Compared to Bro and Snort, the Suricata destroys fewer photons consecutively and detects more malicious messages.
Keywords:*Cloud computing; Anomaly detection; Intrusion detection; Performance analysis; IDS.*

INTRODUCTION:

Since the introduction of electronic devices, the amount of data produced in a single moment has constantly increased and has recently broken through the gigabyte barrier and even entered the terabyte level. This significant growth in data production can be attributed to the steady increase in the number of electronic devices. Businesses in a wide variety of fields are able to increase their profit margins by improving how they handle their resource management and conduct commercial transactions over the network. Since the value of the data is rendered meaningless if its confidentiality and integrity are compromised, maintaining the security of big data continues to be a fundamental priority for all of the proposed solutions. One of the most essential and challenging aspects of network security is the capacity to recognize and prevent network breaches with a high level of accuracy and in a shorter length of time than is currently required for prediction. This is one of the most critical aspects of network safety but also one of the most challenging aspects (Sarumi et al., 2020; Sahu et al., 2021). As a consequence of these incursions, the material's accessibility, as well as the privacy and reliability of the companies provided by big data are all put in jeopardy. Certain providers of big data services utilize firewalls to address the issues outlined earlier in this paragraph (Abid & Jemili, 2020).

The solution requiring the slightest effort to deploy is anomaly-based intrusion detection, initially described in a critical paper. This simple approach works well when dealing with numerical data up to a point, but it fails when dealing with high-dimensional data, such time

series or seasonal information, which is common in real-world applications. The order calls to system functions, library decisions, and machine code are examples of event, can be determined by monitoring software and hardware. The events themselves can be used to deduce this knowledge. Infinite loops are likely to develop if a server processes this data. There will probably be underlying structures that need to be considered while analyzing the data received from social media platforms. Seasonal variations most certainly have an impact on costs and sales for retail firms and other businesses. The previous literature that has been written about this concept is fully explained in the next section.

LITERATURE REVIEW:

AUTHORS AND YEAR	METHODOLOGY	FINDINGS
(Wang et al., 2018)	This study examined cloud-deployed big data TDT apps' multi-layered performance. TDT application performance is affected by these factors. This is the first vertical infrastructure, platform, and software analysis. This study identified critical Every one of the cloud tier's (Infrastructure, Applications, and System) attributes and measurements to establish their connections.	The proposed analysis was tested using Twitter datasets.
(Khan et al., 2022)	This study focused on three network monitoring solutions that immediately detect and respond to cyberattacks. This helps define the paths for tracking networks and investigation into cybersecurity. In order to further comprehend the impacts of intrusions on the internet and track them using instruments like cacti, weather-map, and smocking, this research looked at network tracking methodologies that may be deployed for aiding server and networked equipment.	Unsupervised networks of neurons, artificial intelligence, and the next-generation barrier, and IDS to identify assaults on private networks.
(Vashishtha et al., 2023)	For systems hosted in the cloud,	The proposed model

	a combination security detection technique combines signatures and anomaly-based techniques to find all threats. The algorithm's rate of identifying is substantial at 92.7% on UNSW-NB15, 85.1% on CICIDS2017, and 99.8% on NSL-KDD dataset.	outperforms other models in comparative analysis.
(Moustafa, 2021)	This paper gives a comprehensive IoT-Fog-Cloud structure that explains the links among the three phases for big data analyses and security-related solutions. It discusses security issues, remedies, and architecture research.	Security risks result from the open architecture of IoT and Cloud technologies and privacy issues for fog technology.

Table 1: Literature review

Even while academic research on programme anomaly detection has been fruitful, thorough understanding and evidence of real-world implementation are still lacking, according to previous literature. Unfortunately, most research on identifying programme anomalies is limited to small-scale offline trace analysis in laboratory settings. It is incredibly simple to produce irregularities due to the accessibility of data and computer capability. Finding pertinent irregularities without sounding too many false alarms is difficult, though.

METHODOLOGY:

During this research, a cloud computing environment was set up by using the Cloud lab testbed, which is made available to the community of computer science researchers by a coalition of institutions and industries. The cloud platform employed was OpenStack because of its widespread adoption, resiliency, user-friendliness, and seamless integration with Cloud lab.

Nine of the servers shown in the table above were utilized in the construction of our cloud lab. For this kind of OpenStack installation, OpenStack mandates using a single server to simultaneously fulfil the role of a controller and a network controller node. A total of eight servers will be used for the compute nodes, each of which will house a virtual machine. The machines that were utilized in the installation of the OpenStack Mitaka cloud platform are displayed in Figure 1. immediately following the completion of the OpenStack installation and configuration utilizing Cloud lab services. This study will have built ten virtual machines that acted like typical tenants but had the potential to be hacked and used in an assault. These machines were designed to act as adversaries. This study also constructed an online web server that utilizes a security system that detects intrusions (Snort, Suricata, or Bro) to examine traffic entering and exiting the web server's network.

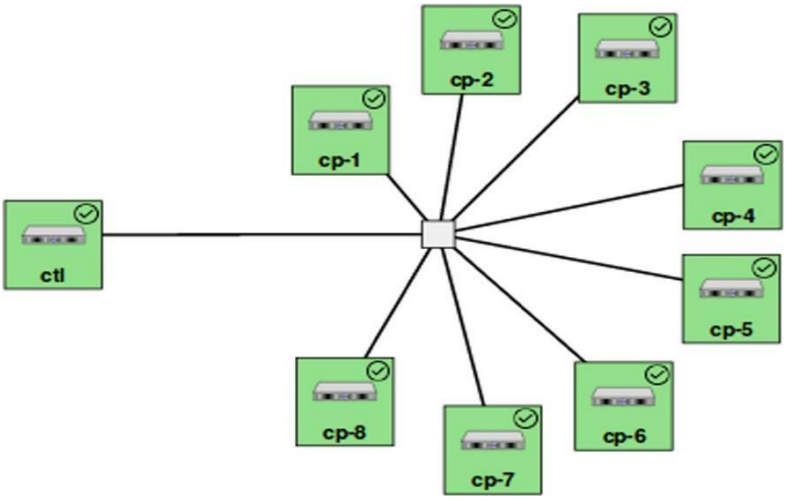

Figure 1: OpenStack Mitaka is installed using administrative units and computing nodes.

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

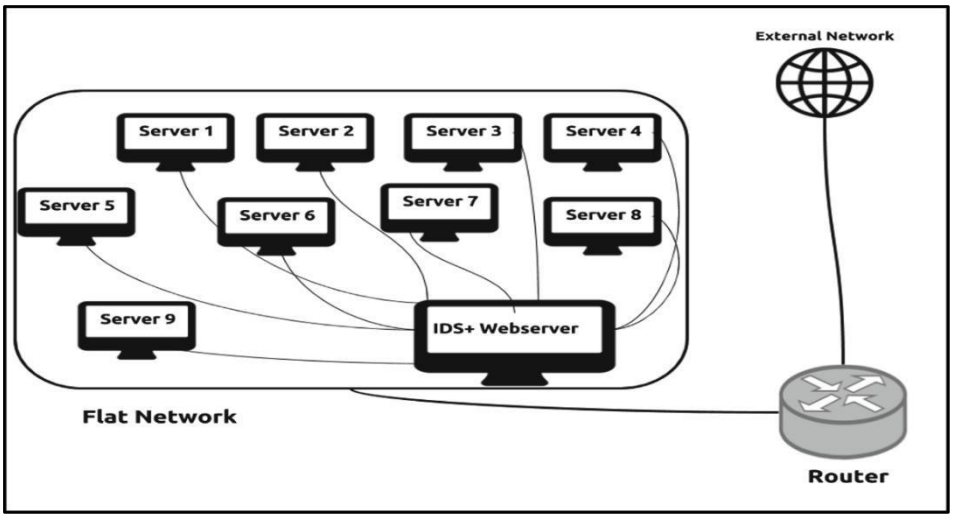

Figure 2: Technical architecture of this study design

RESULTS AND DISCUSSIONS:

After an hour, the DDoS-attacked web server got 9000 HTTP packets. The researchers simply delivered 150 packets per minute to make traffic appear regular. Three intrusion detection systems have evaluated this assault. Every IDS utilized default settings and rules. Figures 3 and 4 show this study's findings. Snort blocked 10% of web server traffic, 900 packets. Suricata dropped only 5% of traffic, or 450 packets. Bro lost 8%. Snort discovers 1970 dangerous packets in the first 15 minutes of the first hour, but its single-threaded nature causes it to reject many more packets in the second half hour.

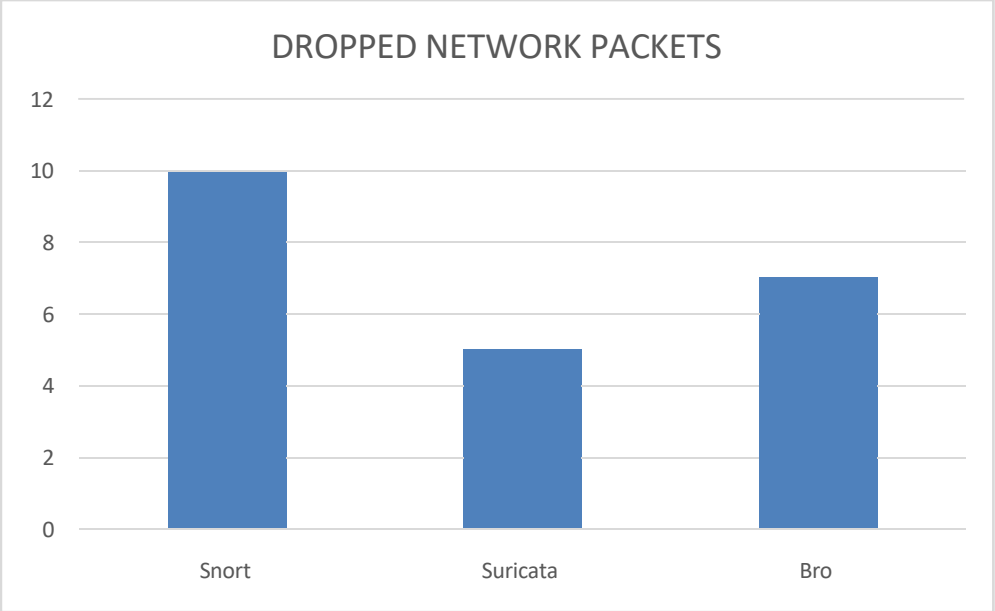

Figure 3: The percentage of packet lost by each IDS.

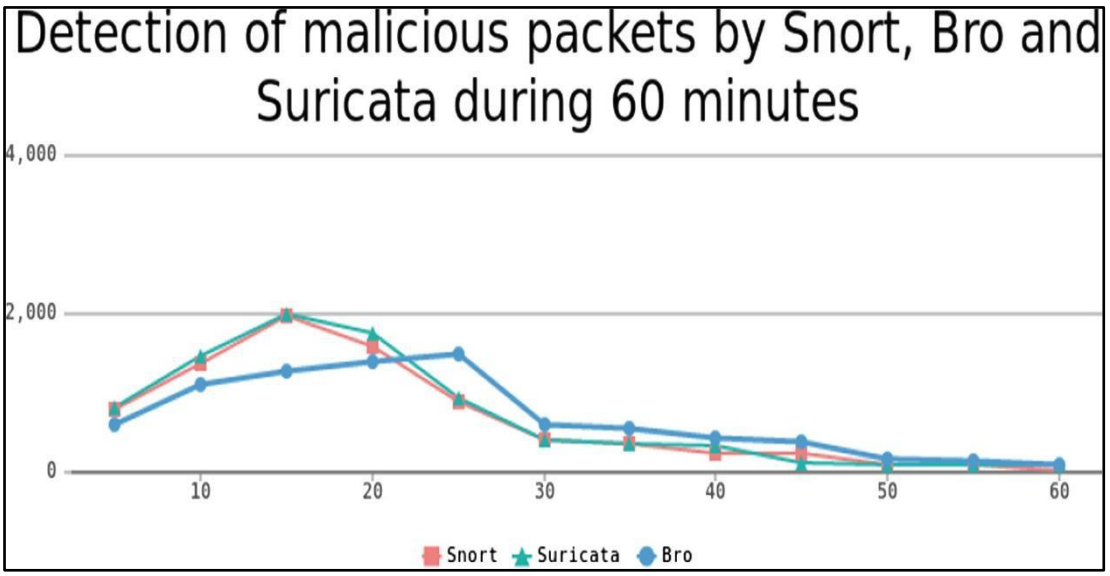

Figure 4: Viruses discovered by Suricata, Bro, and Tabasco in the Last 60 minutes.

Bro found 2000 malicious packets in 15 minutes. After 30 minutes of protesting, the latter released package. Suricata is comparable to Snort, although it loses less communication early on. Snort, suitable for smaller networks, quickly detects malicious conduct. Suricata, like Snort, can substitute. Academic specialists may choose Bro because a community of researchers produces it.

CONCLUSION:

This research discovered that certain IDSs' multiple threads capabilities may assist in preventing packet decreasing that will speed up the processing of dangerous network activity.

Suricata performs better on cloud computing because of multi-threading and many cores. Within 5 minutes, three IDSs discovered the DDoS attack. This research will experiment with more assaults using three of these IDSs throughout time. Each IDS was configured with distinct rules in the present study. This required IoT IDS deployment and testing.

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.