INTRUSION DETECTION SYSTEM IN SOFTWARE-DEFINED NETWORKS USING ML'S XGBOOST ALGORITHM AND OPEN-SOURCE IPS

Olimpjon SHURDI, Ergest ALITE, Aleksandër BIBERAJ and Genci MESI
Faculty of Information Technology, Polytechnic University of Tirana, Albania

---

ABSTRACT

Public cloud has nowadays become one of the most required and used IT platforms—especially during the outbreak of the pandemic caused by the coronavirus. The pandemic period was unfortunately characterised by an increase of cyber-attacks activity. Given the situation, studying and implementing an end-to-end security solution that is based on machine learning's algorithm, called XGBoost, and open-source IPS called snort would have been considered important. The present paper informs about the implementation of an end-to-end cyber security solution to address the protection of public cloud platforms against some different types of security attacks like Distributed Denial of Service (DDoS) attack, Denial of Service (DoS) attack, User to Root (U2R) attack and Remote to Local (R2L) attack, and the results showed its effectiveness when compared to the other methodologies which are using Support Vector Machines (SVM) algorithm.

***Keywords*:** Public Cloud, XGBoost, Distributed Denial of Service (DDoS), Support Vector Machines (SVM)

1. INTRODUCTION

The present paper provides information about cloud computing security due to its great importance and the benefits—especially during the outbreak of the pandemic caused by the coronavirus while labor force working from home with the same efficiency as from the office.

The security of using these platforms has come to the fore due to the high number of users and the high number of already essential services they offer (such as business communication applications Teams, Zoom, etc.). The proposed analysis is based on the study of SECURE (Podlodowski *et al*.,

2019) technique extending it to work for both UDP flood and NTP amplification security attacks— making it appropriate for a comprehensive use. It will also enable us to see changes in the execution performance of this method as a result of its modification by adding protection from additional security attacks.

The present paper reports that the modified technique is one which achieves a defense against more than one type of cyber security attack with the same effectiveness. Here, a network level intrusion detection system called SNORT (Snort 2020) is used. An anomaly detector based on the decision tree and the extreme gradient boosting algorithm (XGBoost) will also be used to analyze abnormal activities (unknown attacks). SNORT is the most effective Intrusion Detection System (IDS). Various Machine Learning (ML) techniques are used for anomaly-based IDSs, but XGBoost is the most widely used anomaly detector based on recent studies. Both of these mechanisms will guarantee the performance and reliability of this method and will ensure the desired result. To extend this method, the attack generator is modified by adding a UDP (UDP Flood) and NTP (NTP Amplification) attack generator to the attack generator. A UDP flood and NTP amplification attack module is also be added to the security attack module. It is noted while analyzing the SECURE technique that this technique protects systems from executing five different types of security attacks, including Denial of Service DoS, Probing, Remote to Local Distance (R2L), User to Root attacks (U2R) and Distributed Denial of Service (DDoS) attacks.

Efforts to use this technique as a defensive infrastructure against other types of malwares such as ransomware attacks, slowloris, etc. will be made in a near future. Here, strategic procedures would be necessary to maintain the effectiveness and reliability.

2. RELATED WORKS

According to a Forbes' report published in 2015, cloud-based security spending is expected to increase by 42%. According to another study, IT security spending had increased to 79.1% by 2015, showing an annual increase of more than 10%. The International Data Corporation (IDC) in 2011 showed that 74.6% of corporate clients rated security as a major challenge. This paper summarizes a number of peer-reviewed articles on Security Threats in Cloud Computing and Preventive Methods. The objective of the research is to understand Cloud components, security issues and risks, along with emerging solutions that can potentially mitigate vulnerabilities in the Cloud. It is a widely accepted fact that Cloud Computing has been a valuable hosting platform since 2008. However, the perception regarding security in the Cloud

is that it needs significant improvements realized at higher levels of adaptation at the company level (Barse *et al*., 2020).

As identified by another research, many of the issues faced by Cloud Computing need to be addressed urgently. The industry has made significant progress in combating threats to Cloud Computing, but there is still more to be done to reach the level of maturity that currently exists with traditional on-premise hosting.

Recent studies on cloud computing security propose ways to protect, but all solutions require a lot of computer resources (Buchanan *et al*., 2016; Khan *et al*., 2016 Gaur *et al*., 2017; Ramachandra *et al*., 2017). Cloud Computing due to its distributed nature, complex architecture and utilized resources poses a unique and serious risk to all actors. Understanding the risk and mitigating the risk appropriately is of critical importance for all the stakeholders. Security must be built at every layer on a Cloud Computing platform incorporating best practices and new technologies to effectively mitigate risk.

Cloud computing allows firms to outsource their entire information technology (IT) process, giving them the opportunity to focus more on their core business to increase their productivity and innovation in customer service. This allows businesses to reduce the heavy cost incurred on IT infrastructure without losing focus on customer needs (Priya *et al*., 2019).

3. **SECURE+, the protection technique in cloud computing**

The present paper aims to create a self-defense and autonomous mechanism against intrusions and cyber attacks (known and unknown) carried out on cloud computing platforms by proposing a technique with self protection and automatic interaction in cloud computing platform called SECURE+. It is an improved version of the SECURE technique by affecting important metrics such as intrusion detection rate (IDR), false positive rate (FPR) and utilization of computer resources (CPU, RAM and bandwidth). SECURE+ will create automatic signatures and provide security against DoS, DDoS, Probing, U2R and R2L security attacks (Kasongo *et al*., 2020). This study is based on the eXtreme Gradient Boosting method (XGBoost), a machine learning algorithm with a set of tree structure-based decisions, which uses a gradient reinforcement framework. The main motivation is the construction of a powerful classification model, which in cooperation with the SNORT system can classify the data entered into the network as accurately as possible, as quickly as possible and with the lowest possible use of available computer resources.

The present paper paper proves that XGBoost is the most suitable method to be used for defensive purposes as a robust classification model could be

built. This will lead to a more accurate intrusion detection system, and a more secure environment to share information on cloud platforms.

The SECURE+ technique helps to detect attacks by a combination of an attack detection system called SNORT and the XGBoost algorithm. SNORT is used to record known attacks on the database that this technique possesses (known attacks). While to detect abnormal activities (unknown attacks) will be used one of the newest techniques based on decision-making tree' algorithm of automatic learning called extreme gradient boosting (XGBoost) (Devan *et al*., 2020). This algorithm allows for the setup of a database which is called the training database, and designs XGBoost to identify and diagnose attacks from incoming network traffic data. The SECURE+ technique will automatically create a new signature and provide security against DoS, DDoS (UDP Flooding and NTP Amplification), probing, U2R and R2L type attacks.

This technique provides a system for detecting intrusion and avoiding computer attacks by way of improving the gradient according to the tree decision technique as in the Figure 1 depicted, automatically, performing an intelligent analysis of packet flow in the network, and being followed by avoidance actions, which are consistent with the decision-making components of intervention detection. The point-to-point detection of intrusion and evasion process is based on three basic applications called Creation of Characteristics, Gradient Enhancement, and Attack Avoidance.

We have chosen to use the gradient boosting algorithm as it is considered the most effective and valuable method in the case of structured data tasks (as is the case with our study).

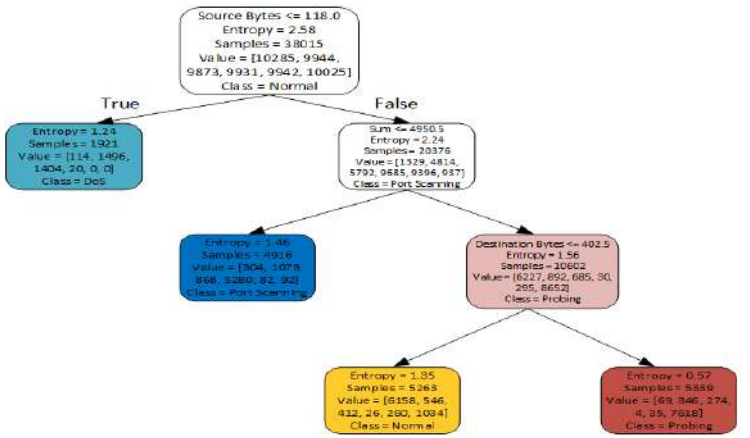

Fig. 16: Decision Tree Algorithm Example for gradient improvement during the task of Intrusion Detection on the network level.

4. **SECURE+ Architecture and Functionality**

The SECURE + technique architecture is in Figure 2 depicted, and its main components are as following:

- Cloud platform users submit their execution requests.
- All user requests are stored by a buffer called the service request handler. This buffer then forwards the workload to the Workload Administrator.
- The workload administrator distributes the workload along with their quality of service (QoS) requests to the Detection Node.

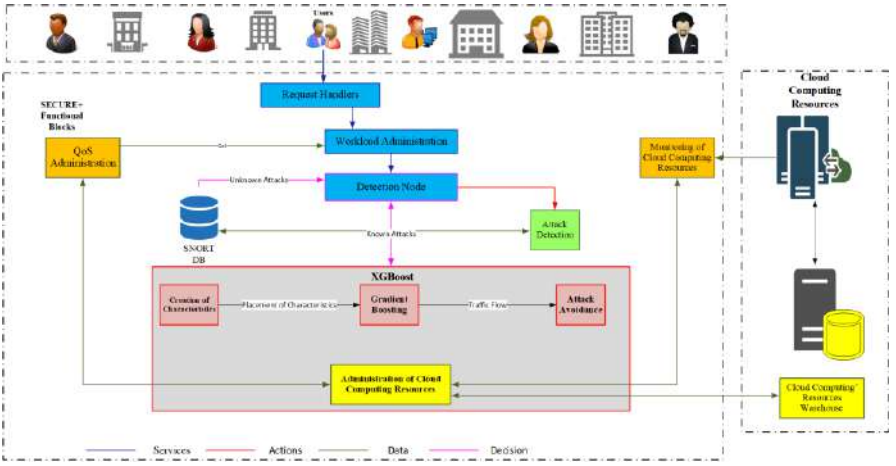

Fig. 2: Architecture of the SECURE+ technique.

- The detection node carries out a two-level defense for protection purposes. Performed by the SNORT intrusion detection application, the first level addresses the known attacks. Performed by the tree decision making (ML) machine learning algorithm, based on the latest XGBoost technique, the Second Level addresses the unknown attacks.
- The Resource Administrator keeps the resource information including the number of CPUs used, the RAM capacity used and the resource numbers. In addition, it stores information about available and reserved resources with respective descriptions (source name, source type, configuration, availability information, and usage information) according to the cloud service provider.
- The autonomous level consists of these three elements: i) creation of characteristics, ii) gradient improvement and, iii) attack avoidance.

- Resource Usage Monitoring which measures the resource usage value during workload execution.
- Cloud Resource Warehouse stores cloud resource configuration.
- Autonomous Attack Detection System Via Machine Learning Technique

The SECURE+ technique considers three steps regarding the autonomous detection and interaction system. The interaction of these subunits such as creation of characteristics, gradient enhancement / increase, and attack avoidance is in Figure 3 described.

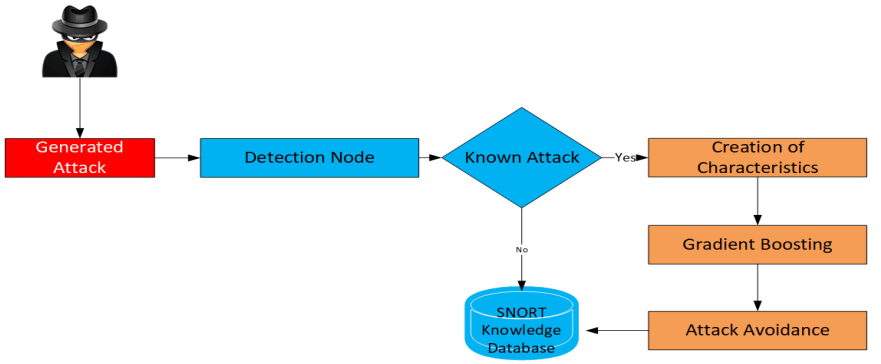

Fig. 3: Interaction of autonomous system's subunits.

The feature creator collects network traffic from the intrusion detection node in case the attack is unknown to snort and calculates the value required by the gradient enhancer for the respective data's flux.

The gradient enhancer applies its pre-built intervention detection models to the flow pattern and passes the result towards the Attack Avoider.

The attack avoider then determines the action to be taken, based on the classification result, and installs the flow rules in the attack detector to prevent the attack if necessary.

Figure 4 depicts the block diagram of the point-to-point function of the proposed security solution.

The detection node collects network traffic data and if they belong to unknown attacks, they are taken over by the feature creator. After acquisition, characteristics are created for each flow as summarized in the block diagram in Figure 5.

In this way common characteristics are generated for each stream, enclosing each incoming stream in the detection node, and using the block diagram shown in Figure 6, e.g., the average flow time and the total number of packets in a transaction are created with an initial switch overflow input.

Consequently, specific flow characteristics, e.g., the duration of the flow and the calculation of packets from source to destination are obtained by passing overflow inputs.

As we look at the stream inputs, the feature vector created for one stream is immediately sent to the Gradient Enhancer, without waiting for the feature creation for the other flows inputs to finish.

- The characteristics creator also draws flows matching fields, such as the source IP and MAC addresses, and the physical gateway from which the packet comes from. Attack avoidance uses these features as a conclusion. Common features include Mean, Stddev, Sum, TnP_PSrcIP, TnP_PDstIP, and TnP_Per_Dport. Hash groups are used to store unique source and destination IPs and destination port numbers. A list is implemented to save the duration of the input stream. As it surrounds the input stream, the number of packets in the input stream is added to the total packet calculation. The input stream duration is added to the duration list. Source IPs, destination IPs and destination port numbers are added to the respective hash group. The flow bit count is added to a hash map. The keys in this map consist of the source IP, the source port, the destination IP, and the destination port for the TCP and UDP packets.
- For ICMP packets, the keys consist of the source IP and the destination IP, unless in this case they do not have port numbers. This map is later used to obtain reverse flow statistics. Once the detection node is enclosed, the main characteristics are calculated using the total packet count, hash groups and duration list.
- The gradient enhancer takes feature vectors from the characteristics creator one by one and classifies them using its prefabricated intrusion detection model. If the result of the classification is any type of attack, then attack avoidant takes the type of detected attack and source identifiers such as. Source IP and source MAC address. The automated learning model used is dynamically updated by including new data learned about the same existing types of attacks as soon as they are detected. The gradient-enhanced model construction is a multi-class classification model, which is formed using learned data that exists in different types of attacks in addition to normal traffic.

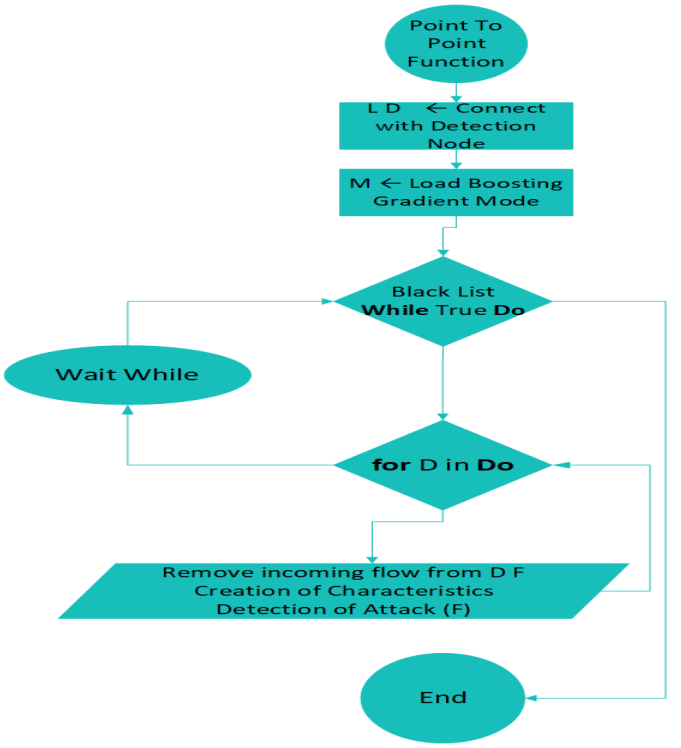

Fig. 4: Block Diagram of Point-to-Point' Function.

4.1. Description of Research Methodology

We designed the experiment scenarios to observe and obtain the measurement results. In the present investigation the performance between two rigorously repetitive and quantitative techniques is compared. The experiments are based on a test environment which compares the SECURE technique with the SECURE+technique.

This environment will have a 100Gbps bandwidth network and several different types of harmful traffic.

Harmful traffic types are selected because the default rules can be applied simultaneously to SECURE and SECURE+. Moreover, these are the most common types and cover a much larger number of attack's types.

The performed experiments will compare the performance of the two techniques by measuring the False Positive Rate (FPR), the Intrusion Detection Rate (IDR), the execution time, as well as the percentage of CPU, the use of RAM memory and the percentage of lost packets on the network.

Normal network traffic for conducting experiments is realized using an open-source network traffic generator called Hping3 (Hping, 2020). This application will generate network traffic up to 20 Gbps. Malicious traffic is generated using the following applications:

- Metasploit for DoS attacks,
- NMAP for Probing attacks,
- Hydra for R2L attacks,
- NetCat for L2R attacks
- DDoSIM for DDoS attacks.

Legitimate and malicious network traffic is generated as combined traffic and then entered into the detection node.

4.2 Realization of Experimental Work

Experiments were carried out to show the advantages of the SECURE+ technique versus the SECURE one using virtual machines with the same computer resources.

First, the experiment was carried out so that real-time observes the performance of the SECURE+ and SECURE techniques by processing a legitimate 10Gbps traffic from the legitimate traffic generator (Hping3) for comparison purposes.

1,470-byte packets for TCP, UDP and ICMP protocols were used for the accurate results of the experiment. We injected these packets in both techniques with a network speed of 10Gbps. The experiment is based on the logic grid diagram as in Figure 5 depicted.

We have installed each technique separately in identical virtual machines with the same parameters of computer resources and the same rules for the snort application. We used an application called Network Performance Monitor by Solarwinds (SolarWinds 2020), which records and measures CPU, memory, and network usage. In addition to this application, we have used several other applications such as Metasploit structure, Snort logs, nmap etc.to record and measure the features.

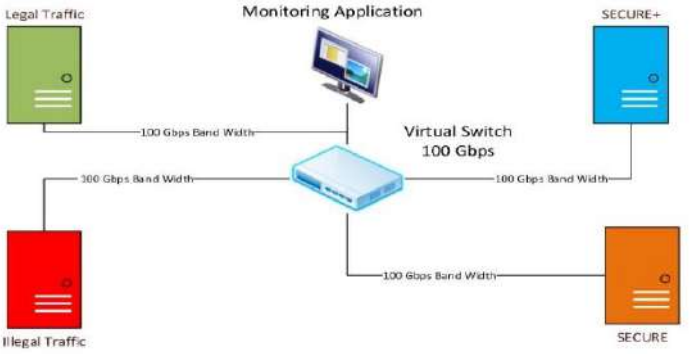

Fig. 5: Presentation of the experimental environment network.

5. EXPERIMENTAL RESULTS

The experiments were carried out in cycles lasting 8 hours each. In total we have performed 10 cycles of tests, and we have achieved a total duration of 80 hours of experiments, this to increase the reliability of our results. The following packets are injected as basic traffic ranging from 1Gbps to 10Gbps as follows:

- 1,000,000 UDP packets at a speed of 500 packets / second, each packet size consists of 1,470 bytes.
- 1,000,000 TCP packets at a speed of 500 packets / second, each packet size consists of 1,470 bytes.
- 1,000,000 ICMP packets at a speed of 500 packets / second, each packet size consists of 1,470 bytes.

Result analysis

The way we chose to inject the packets was the one with normal traffic, specifying the number of packets per second and the total number of packets.

The results showed that the use of CPUs in the SECURE+ technique was lower compared to that of the SECURE technique measured during the processing of the same network traffic of 10Gbps.

The Figure 6 gives the average CPU usage score for both techniques during the 80 hours of testing.

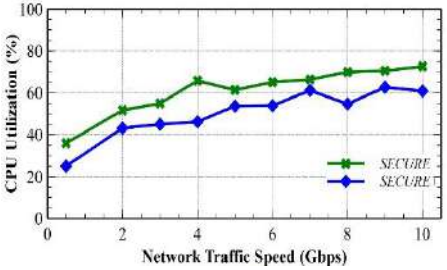

Fig. 6: Average CPU usage.

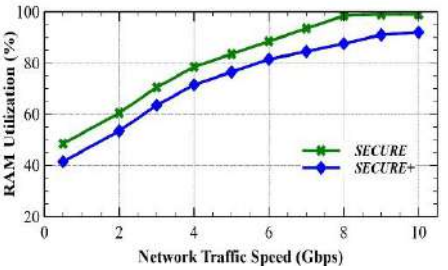

Fig. 7: Average RAM usage.

The collected performance data tells us that the memory usage in the SECURE+ technique is lower than that of the SECURE technique.

It is clear that the average memory usage in the case of the SECURE+ technique increases from 19GB when performing tests at 1Gbps and continues to increase at a variable rate up to a maximum of 30GB when performing tests with 10Gbps network speed. Graphically this usage is in Figure 7 depicted.

Processing packages in the SECURE+ technique are faster than processing packages in the SECURE technique as per Figure 8.

In other words, for the same amount of UDP, TCP and ICMP packets (1,000,000 packets) injected in both techniques over a period of 80 hours (10 test cycles of 8 hours each) we noticed that the SECURE+ technique showed an improved performance versus that of SECURE.

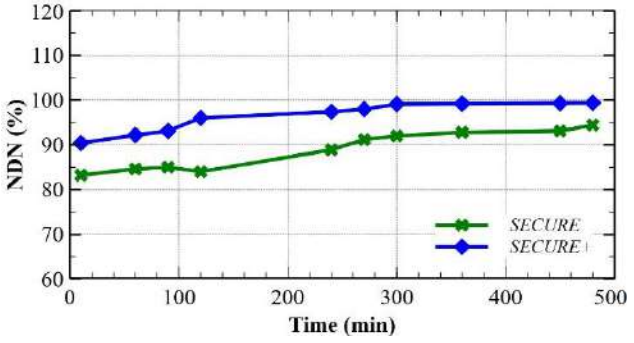

Fig. 8: Average processing speed (number of packets per unit time) for both techniques.

To compare the performance of SECURE and SECURE+ techniques, two most important metrics such as intrusion detection rate (IDR) and false positive rate (FPR) were investigated as they are also the most important criteria for evaluating algorithms and techniques.

The results obtained for these metrics were realized by generating attacks through Metasploit applications for DoS attacks, NMAP for probing attacks,

Hydra for R2L attacks, NetCat for L2R attacks and DDoSIM for DDoS attacks.

Figure 9 graphically shows the results obtained during the experiments performed for the mean values of false positive rate (FPR) for all categories of attacks (4 types of attacks, DoS, R2L, Probing and DDoS).

Mostly, the algorithms that have a low value of the IDR metric are not considered at all, and are not used, no matter how high the value of the FPR.

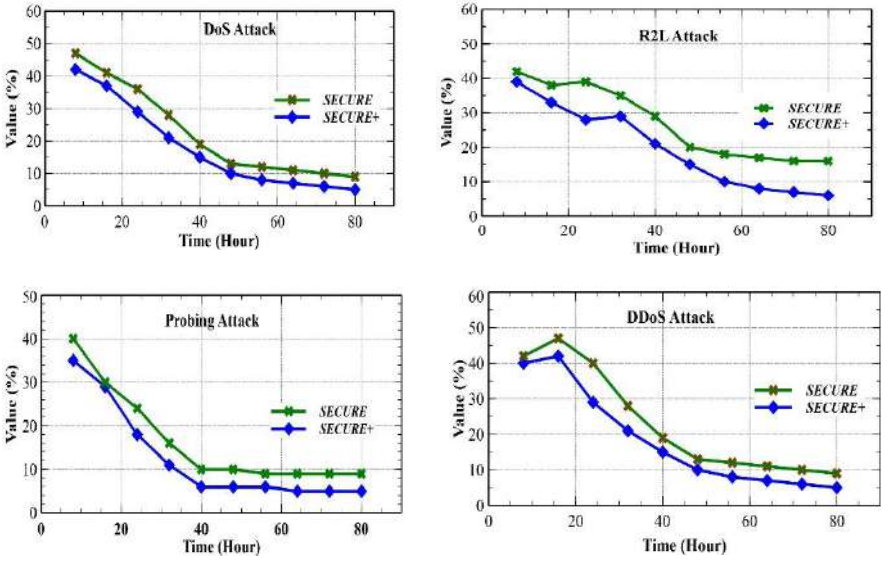

Fig. 9: Results of experiments for FPR for all types of attacks.

The Figure 10 depicts that the intrusion detection rate (IDR) increases with time. In this case we performed the same experiment divided into 10 cycles of 8 hours each.

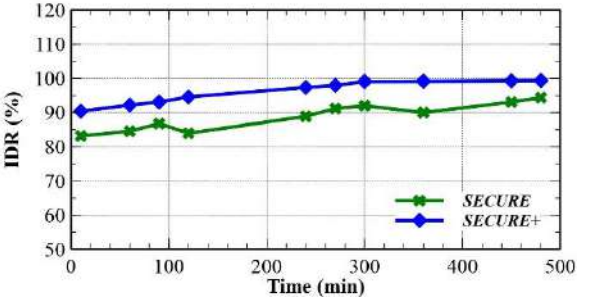

Fig. 10: Results of experiments for FPR for all types of attacks first and last cycles.

6. CONCLUSIONS AND FUTURE WORK

In this paper we have proposed a technique with self-protection and automatic interaction in cloud computing platforms called SECURE+ is proposed. This self-defence and automatic technique detect intrusions and attacks (known and unknown) carried out on Cloud Computing platforms. The SECURE+ technique is built to detect attacks by a combination of an attack detection system called SNORT and a machine learning algorithm, called XGBoost. It functions as a single system of interventions where a series of snort rules operate in parallel with the logic applied by the tree-based automated learning algorithm XGBoost.

SECURE+ protects cloud computing platforms from five different types of security attacks including DoS, DDoS (also UDP Flooding and NTP amplification), Probing, U2R, and R2L attacks. Furthermore, we have tested the performance of the SECURE+ technique in terms of intrusion detection rate, execution time, false positive rate and use of computer resources.

The present paper compares two intervention detection techniques, the SECURE+ and SECURE are. Both techniques turned out to be efficient and high-performance detection systems. The results showed that the SECURE+ technique is more effective and uses less computer resources compared to the SECURE technique as it uses approximately 10% less processing resources (CPU) and about 7% less RAM. In addition, the SECURE+ technique processes a larger number of packets about 30% more packages per second compared to the SECURE technique.

The SECURE+ technique has a lower packet loss rate than the SECURE technique. In both techniques it was observed that the operating system was responsible for packet losses in case of increased traffic from 2Gbps to 10Gbps. At these network speeds, memory buffers were completely occupied by not being able to read packets within these buffers, so both techniques require more RAM in the case of high network capacities ranging from 2Gbps to 10Gbps. This phenomenon does not exist for low network speeds from 100Mbps to 1Gbps.

The results of the experiments showed that the SECURE+ technique has a false positive rate about 4% lower than the SECURE technique.

The other very important metric investigated was the intrusion detection rate, as it determines the efficiency of a detection system. Results showed that the value of IDR increased with increasing time reaching the maximum value of 98.8% during the 40th hour.

Considering the proposed and studied technique, realization and results obtained, we conclude that the following works can be based on the future for such technique:

- Realization of the SECURE+ technique in a real Cloud environment.

- Improved SECURE+ for identifying and preventing day-zero attacks.
- Improvement of SECURE+ to identify the rate of breach of service level agreement (SLA).
- Improved SECURE+ to work with some other parameters such as energy efficiency, scalability, etc.

REFERENCES

Barse Y, Agrawal D.2020. BotNet Detection for Network Traffic using Ensemble Machine Learning Method. *International Journal of Innovative Technology and Exploring Engineering* (IJITEE), **10(1):** 201-204. ISSN: 2278-3075. 100.1/ijitee.A81221110120 DOI: 10.35940/ijitee.A8122.1110120.

Buchanan J, Flandrin B, Macfarlane F, Graves R. 2016. A methodology to evaluate rate-based intrusion prevention system against distributed denial of service.

Devan P, Khare N. 2020. An efficient XGBoost–DNN-based classification model for network intrusion detection system. *Neural Computing and Applications*, **32**:12499–12514. SpringerLink. https://doi.org/10.1007/s00521-020-04708-x.

Hping. 2020. http://www.hping.org/.

Kasongo S.M, Sun Y.2020. Performance Analysis of Intrusion Detection Systems Using a Feature Selection Method on the UNSW-NB15 Dataset. *Journal of Big Data*, **7:** 105 https://doi.org/10.1186/s40537-020-00379-6.

Khan N, Al-Yasiri A .2016. Identifying cloud security threats to strengthen cloud computing adoption framework. *Procedia Computer Science*, **94**: 485-490.

Podlodowski L, Kozlowski M. 2019. Application of XGBoost to the cyber-security problem of detecting suspicious network traffic events. 2019 IEEE International Conference on Big Data (Big Data), 10.1109/BigData47090.2019.9006586.

Priya MSS, Sahu BK, Kumar B, Yadav M. 2019. Network intrusion detection system using XG Boost. *International Journal of Engineering and Advanced Technology* (IJEAT), **9(1):** 4070-4073. ISSN: 2249 – 8958**.** 10.35940/ijeat.A1307.109119.

Ramachandra G, Iftikhar M, Khan FA. 2017. A comprehensive survey on security in Cloud Computing. *Procedia Computer Science*, **110**: 465-472. https://doi.org/10.1016/j.procs.2017.06.124. Elsevier B.V.

Roundup of Cloud Computing Forecasts and Market Estimates. 2015. http://www.forbes.com/sites/louiscolumbus/2015/01/24/roundup-of-cloud-computing-forecasts-and-market-estimates-2015/#56c0b0f0740.

Snort. 2020. https://www.snort.org/

SolarWinds 2020. Network Performance Monitor, https://www.solarwinds.com/network-performance-monitor