International e-Conference on New Horizons And Multidisciplinary Applications In Science And Technology
In Association withInternational Journal of Scientific Research in Science and Technology
Volume 9 | Issue 6 | Print ISSN: 2395-6011 | Online ISSN: 2395-602X (www.ijsrst.com)

Using Machine Learning Techniques to Detect Distributed Denial of Service Attacks

Ms. Varkha K. Jewani(Ms. Pragati V. Thawani)[1], Dr. Prafulla E. Ajmire[2], Ms. Geeta N. Brijwani[1]

[1]Assistant Professor, Department of Computer Science, K.C College, Church gate, Sant Gadge Baba Amravati University, Maharashtra, India

[2]Head & Associate Professor, Department of Computer Science & Application, G S Science, Arts & Commerce College, Khamgaon, Maharashtra, India

ABSTRACT

Machine learning (ML) is used for network intrusion detection because it is predictable after training with relevant data. ML provides a great way to detect new and unknown attacks. There are many types of network intrusion attacks; however, this document focuses on distributed denial of service (DDoS). DDoS attacks are the most destructive attacks, which will disrupt the safe operation of basic services provided by different organizations in the Internet community. These attacks are becoming more and more complex, and the number is expected to increase, which makes detecting and combating these threats challenging. Therefore, an advanced intrusion detection system (IDS) is needed to identify and recognize abnormal behavior of Internet traffic. This research combines well-known clustering methods such as Naive Bayes, Multilayer Perceptron (MLP), and SVM, uses decision trees and various classification algorithms, to detect DDOS attacks.

Keywords: Classification, Distributed Denial of Service, Machine Learning,

I. INTRODUCTION

With the expansion of computer networks (especially the Internet), a variety of cyber-attacks have appeared. An international ransomware virus called Wannacry recently stopped internet services in approximately 156 countries. Based on Kaspersky Lab's full Q4 results, botnet-assisted attacks target assets in nearly 69 countries / regions. The last quarter also saw the largest DDoS-based botnet attack lasting approximately 15.5 days and 371 hours. Hackers or shady hackers continue to create new forms of multi- layered DDoS attacks, mainly on the OSI network and the application layer [1]. This type of attack uses spoofed IP addresses to hide source detection and carry out large- scale attacks. These attacks are quite large, because the attack traffic absolutely consumes the network spectrum at peak times, thereby reducing legitimate data packets.

Ironically, the victims were government entities, financial companies, national defense forces, and military institutions. Famous websites such as Facebook, Twitter, and WikiLeaks have become victims of DDoS. They have also observed that the interruption of routine maintenance leads to financial failures, exhaustion of services, and inaccessibility [2]. This paper focus on different machine learning identification methods, such as SVM, Naive Bayes, and decision trees, to detect and analyze different forms of these attacks, including Smurf, UDP flood, and HTTP flood.

In Network Intrusion Detection System (NIDS) research, there are three detection methods: misuse or signature- based, anomaly-based, and hybrid-based detection methods. Signature-based or misuse methods primarily detect known intrusion attacks, while anomaly-based methods detect new or unknown intrusion attacks. The hybrid base can detect known and unknown intrusion attacks [3].

Machine learning (ML) technology learns patterns from past data and makes predictions on current data. Because ML recognizes patterns, rather than specific signatures, it can be used in hybrid-based methods that can detect small changes in known attacks. New attacks are constantly occurring, so it is very important that NIDS can adapt to changes to detect known and unknown attacks. There are many types of attacks on NIDS. However, this article focuses on distributed denial of service (DDoS) attacks [4]. DDoS is very similar to a denial-of-service attack. The difference is that the latter has a single source of attack, while the former has multiple sources of attack. Due to the total consumption of network resources by these attacks, both types of attacks will result in the inability to access network resources. The challenge of an effective NIDS is to have a high accuracy rate, a low false alarm rate. These are some of the main indicators emphasized by the current NIDS research [5].

II. TYPES OF ATTACKS

DDOS Attack – If multiple applications (usually one or more application servers) flood the capacity or infrastructure of the target network, a distributed denial of service (DDoS) attack can occur. Figure 1 below shows an attack that is usually the result of multiple infected systems (such as botnets) flooding the target network with traffic [7, 8].

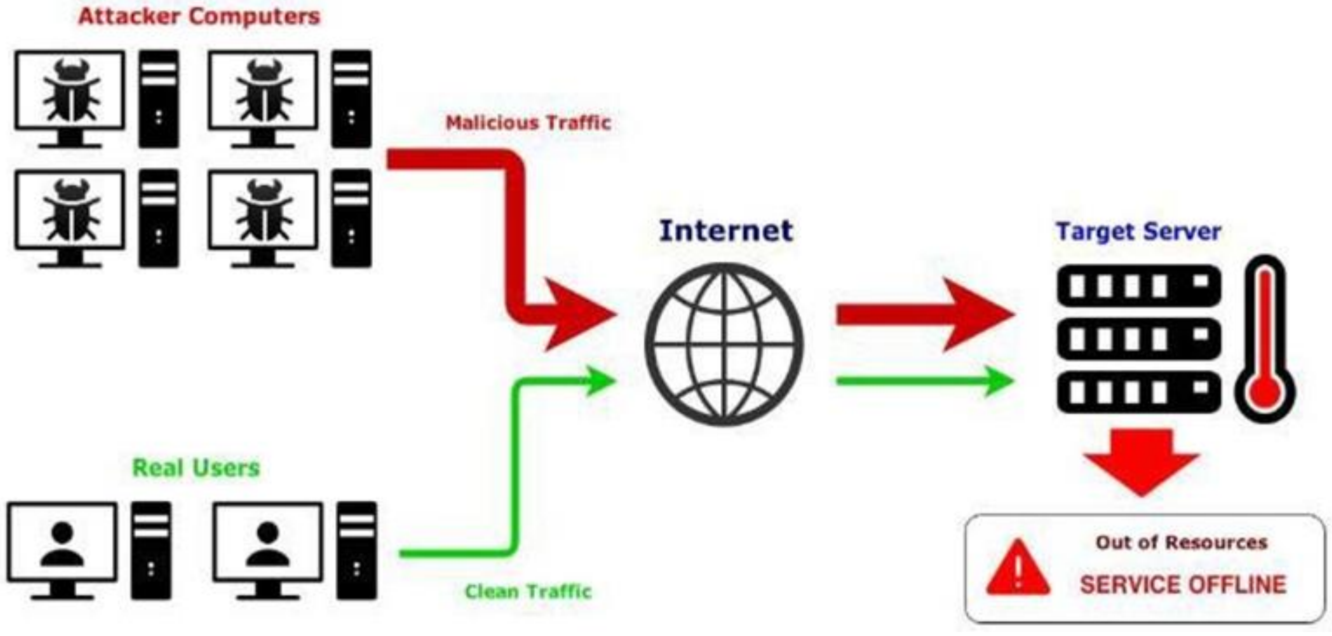

Figure 1: DDOS Attack

UDP Flood - UDP flooding is a Denial of Service (DoS) attack in which the attacker uses an IP packet composed of User Datagram Protocol (UDP) packets to attack and bypass the host's random port. Figure 2 below shows that the host is looking for applications related to certain datagrams in the entire attack pattern. If it is not detected, the host will send a "target unreachable" packet to the sender. The result of this flood bombing is that the network is flooded and therefore does not respond to legal traffic [8].

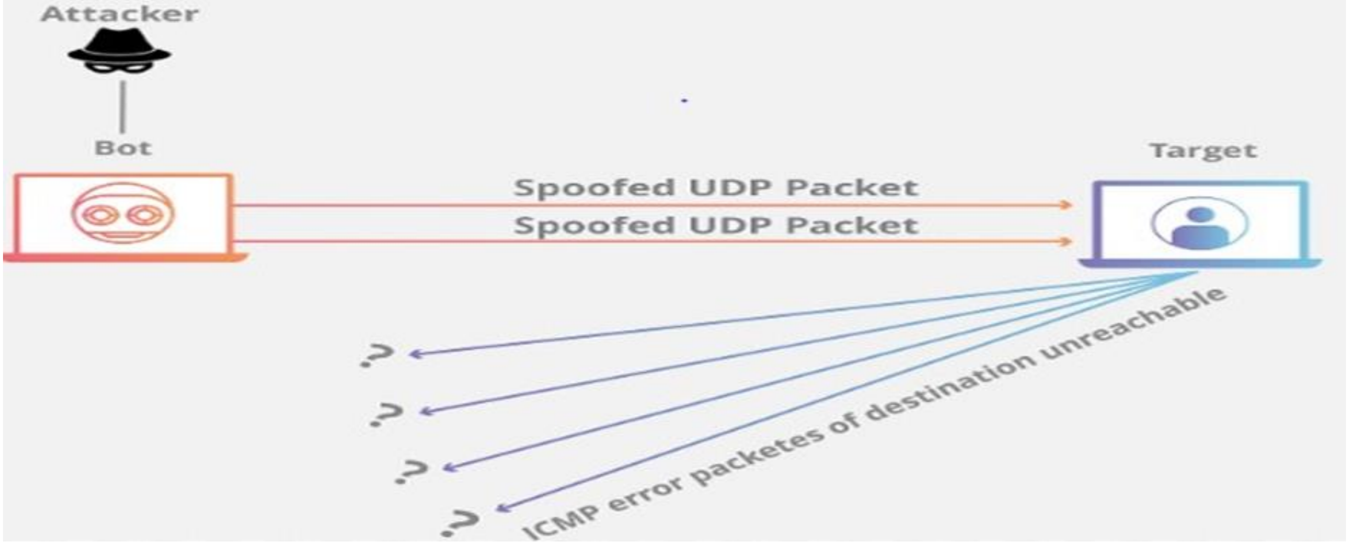

Figure 2: UDP Flood

ICMP (PING) Flood - Ping flood, is recognized as ICMP flood, is a well-known Denial of Service (DoS) assault where attacker powers a casualty's gadget down with flooding it with demands for ICMP reverberation, likewise called as pings. The figure3 will clarifies ICMP flood assault, this attack incorporates overpowering the casualty's organization according to popular demand parcels, understanding the framework will respond with similarly however many answer bundles as could reasonably be expected. Record types to get an objective down for ICMP demands additionally utilize custom programming or code, such as hping and scapy [9].

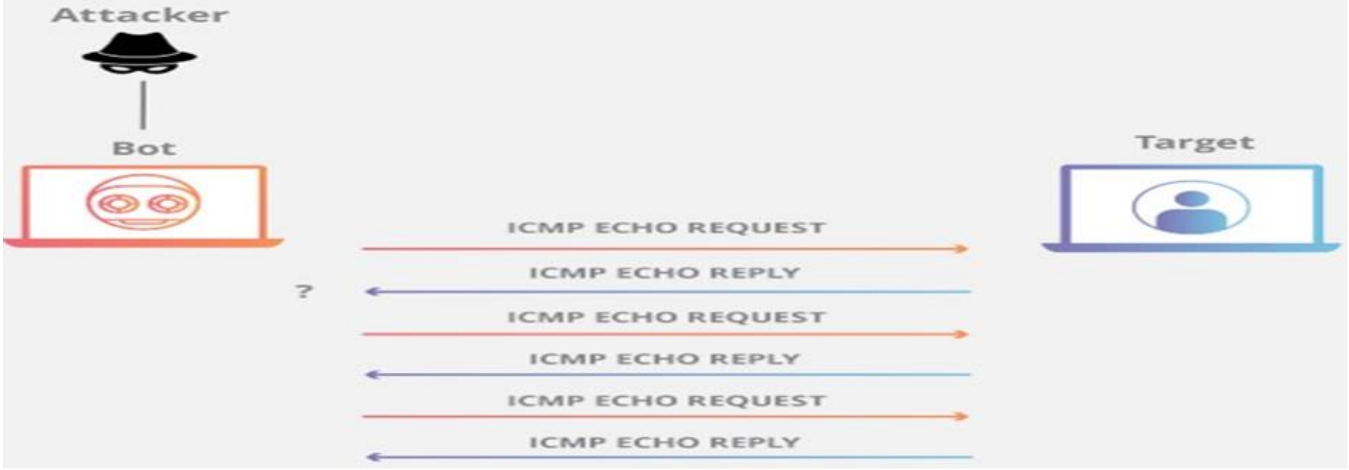

Figure 3: Icmp (ping) Flood

SMURF Attack This is additionally a one of the ddos attack wherein enormous gatherings of Internet Control Message Protocol (ICMP) bundles for the most part utilizing satirize source IP of the casualty are communicated over an IP broadcast address to a PC organization. The beneath figure4 shows naturally, numerous gadgets on an organization will answer it by giving a reaction to the source IP address. On the off chance that the number of frameworks over the organization getting, reacting to such bundle is very high, so traffic can overpower the assailant's PC [9.10].

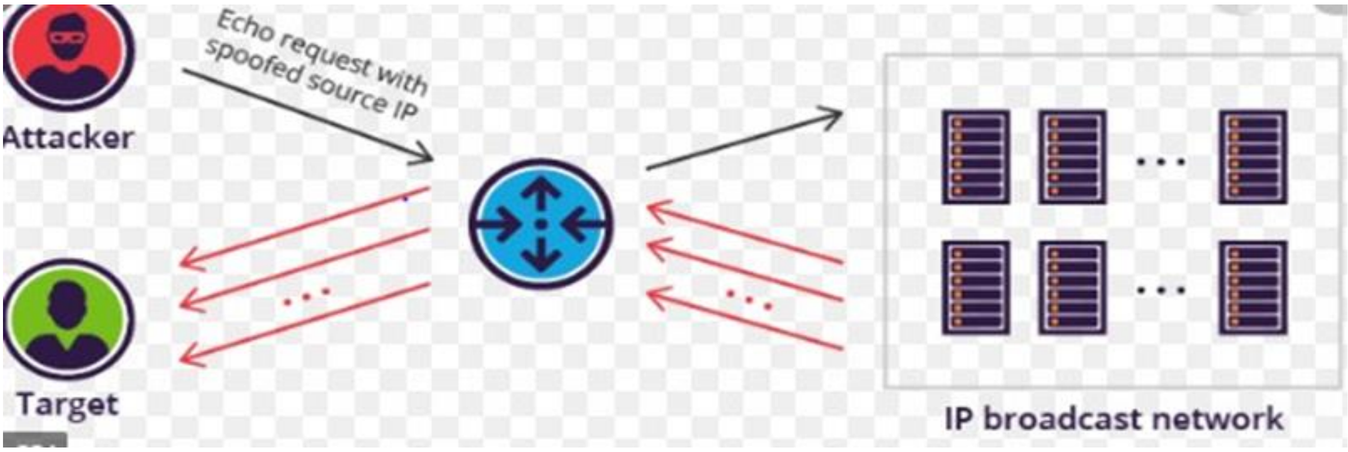

Figure 4: Smurf Attack

Http Flood Attack -A HTTP flood is a refusal of- administration dispersed volumetric (DDoS) assault, it is displayed in the figure5, and it is worked to overburden a chose worker with HTTP demands. At the point when the objective has likewise been loaded up with questions and can't respond to customary traffic, there will be forswearing of-administration for explicit solicitations from real clients [9, 10].

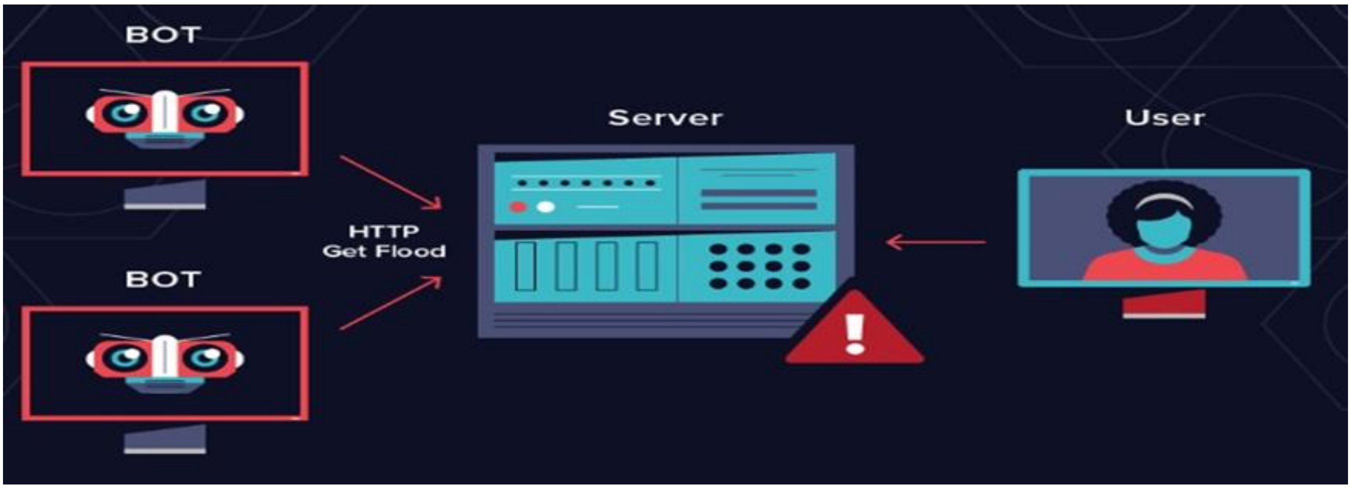

Figure 5: Http flood Attack

III. MACHINE LEARNING METHODS RELATED TO DDOS ATTACK DETECTION

Machine learning (ML) strategies study the sample from beyond information and make predictions for modern-day information. Since ML recognizes patterns, as opposed to specific signatures, it could be used for hybrid-primarily based totally technique that could hit upon small versions from regarded assaults. New assaults are constantly being created, consequently it's far crucial that NIDS be capable of adapt to adjustments to hit upon each regarded and unknown assaults [11].

ML can be partitioned into three fundamental sorts: supervised, unsupervised and semi-supervised. Supervised calculations expect information to be marked, then, at that point dependent on the name, they can characterize the information as indicated by an unmistakable example for each class or name. Solo calculations can utilize information with no naming. This sort of calculation bunches the information into group(s) with comparative qualities. Semi- supervised calculations use information that are somewhat marked. The ML types and calculation are displayed in Fig. 6. It has been tracked down that administered calculations function admirably in IDS with recently known assaults, while solo calculation are more hearty with both known and obscure assaults [11,12].

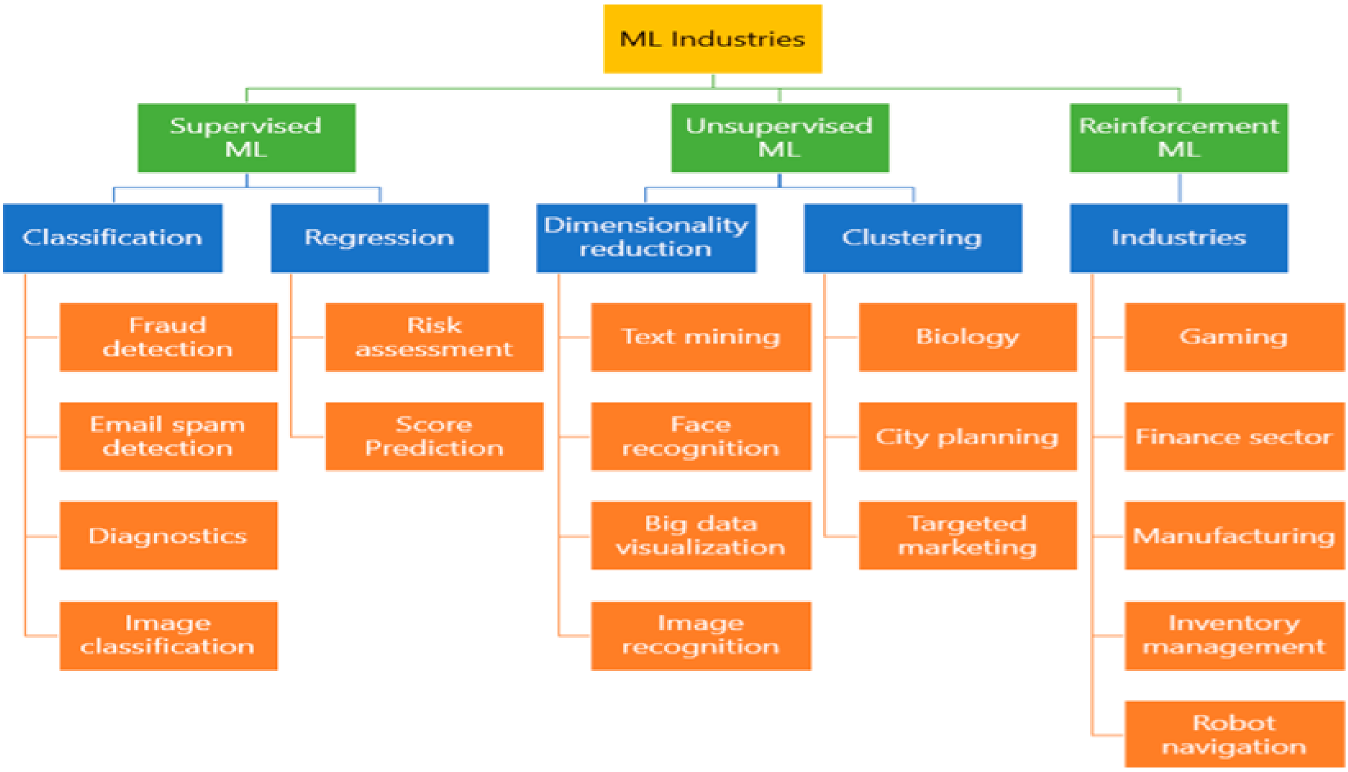

Figure 6: Classification of ML Algorithms

Signature based IDS is a human based activity, including numerous long periods of testing, creating and conveying the mark and making new mark for obscure assaults as well. So giving a less human based framework becomes essential. Machine Learning dialects inferred abnormality based IDS offers an answer for this issue, assisting with fusing a system which can gain from information and anticipate obscure details data on educated information [13, 14].

A. **Naïve bayes.** -Naive Bayes is cantered around the Bayesian arrangement model. Building up classifiers is a simple and easiest strategy: models which gives class names to give cases, characterized as the vectors of highlighting esteems, where the classes marks will be determined among certain limited set [14].

B. **Support Vector Machine**- SVM makes order and relapse utilizing the directed strategy for learning .Based on a gathering of trained models, every one of which is set apart as strategies are partitioned into two arrangements, a SVM calculation makes a plan which predicts that the new model will in general fall into one among the two [14].

C. **Decision Trees** - One of the essential methods utilized in AI and information mining is the decision tree. It is additionally used as a prescient model where discoveries with respect to an article are planned to suspicions about the ideal worth of the item. A decision tree might be utilized in the choice information examination to outwardly and expressly demonstrate dynamic. The informational index is contemplated and developed in this strategy. Therefore, if the new information component is given for characterization, the earlier dataset will group it properly [15, 16].

D. **Artificial Neural Network** -The expression "Artificial Neural Network" is gotten from Biological neural organizations that foster the construction of a human mind. Like the human cerebrum that has neurons interconnected to each other, artificial neural organizations additionally have neurons that are interconnected to each other in different layers of the organizations. These neurons are known as nodes [16].

E. **K-means clustering** - It is a bunching procedure broadly used to parcel an assortment of information in bunches k naturally. The K-implies grouping calculation works by picking k beginning bunch places in an informational collection and afterward refining them recursively as portrays. .Each model will be dispensed to its closest bunch center. It refreshes the mean of its part cases to every one of the bunch places. The calculation merges when the allotment of occasions to groups doesn't adjust further [15,16].

F. **Logistic Regression** - This calculation utilizes a regression model to track down the best-fitting model that portrays a reliant variable dependent on a bunch of autonomous factors. The results of the reliant variable comprise of just two potential qualities: valid or bogus. Along these lines it is appropriate for parallel characterizations [14, 15].

G. **Boosted Trees (BT)** - This calculation depends on choice tree with the expansion of a boosting technique. All things considered, of building one enormous tree, various little trees are constructed. Then, at that point the consequence of every little tree is added, with a weighted worth, to get a last prescient result [15, 16].

H. **Random Forest** - This calculation is like BT, where various little trees are constructed. Be that as it may, it contrasts in the manner in which it works out the last prescient result. Rather than utilizing a boosting strategy, it utilizes a packing technique. This strategy utilizes the mean of the singular little trees to get the last prescient result. This classifier is observed to be quick and proficient with enormous datasets [15, 16].

IV. CONCLUSION

It is concluded after a detailed analysis that web attacks are risky and that IDS / IPS may not tackle the new attacks that affect the networks. Machine learning approaches play a critical role in gaining exposure to the intensity of the assault and thereby making enterprises take suitable measures to limit certain attacks.

V. ACKNOWLEDGMENT

The author is very much thankful to Dr. P.E Ajmire sir for all his guidance and ICMA2021 for giving this opportunity to publish research paper for this particular conference and all the members of conference team.

VI. REFERENCES

[1] . Nisioti A, Mylonas A, Yoo PD, Member S, Katos V. From Intrusion Detection to Attacker Attribution: A Comprehensive Survey of Unsupervised Methods. IEEE Commun Surv Tutorials. 2018; PP(c):1.

[2] . Ahmad I, Basheri M, Iqbal MJ, Rahim A. Performance Comparison of Support Vector Machine, Random Forest, and Extreme Learning Machine for Intrusion Detection. 2018; 33789–95.

[3] . Zou X, Feng Y, Li H, Algorithm MC, Hamid IRA, Syafiqah N. Performances of Machine Learning Algorithms for Binary Classification of Network Anomaly Detection System Performances of Machine Learning Algorithms for Binary Classification of Network Anomaly Detection System. 2018;

[4] . Biswas SK. Intrusion Detection Using Machine Learning: A Comparison Study. 2018;118(19):101–14.

[5] . Ahmad B, Jian W, Ali ZA. Role of Machine Learning and Data Mining in Internet Security : Standing State with Future Directions. 2018.

[6] . Ramírez-Gallego S, Krawczyk B, García S, Woźniak M, Herrera F. A survey on data preprocessing for data stream mining: Current status and future directions.Neurocomputing. 2017;239:39–57

[7] . Informatics S, Science C. An Ensemble Approach Based On Decision Tree And Bayesian Network For Intrusion Detection. Comput Sci Ser. 2017; 15:82–91.

[8] . Khan, I.U.; Shahzad, M.U.; Hassan, M.A. Internet of Things (IoTs): Applications in Home Automation. IJSEAT 2017, 5, 79–84.

[9] . Shams R, Mercer RE. Supervised classification of spam emails with natural language stylometry. Neural Comput Appl. 2016; 27(8):2315–31.

[10] . Yogeswara Reddy B, Srinivas Rao J, Suresh Kumar T, Nagarjuna A, International Journal of Innovative Technology and Exploring Engineering, Vol.8, No. 11, 2019, pp: 1194-1198.

[11] . Ch. Mallikarjuna Rao, G. Ramesh, Madhavi, K., "Feature Selection Based Supervised Learning Method for Network Intrusion Detection", International Journal of Recent Technology and Engineering (IJRTE), ISSN: 2277- 3878, Volume-8, Issue-1, and May 2019.

[12] . Thirupathi, N., Madhavi K., Ramesh G., Sowmya Priya, K. "Data Storage in Cloud Using Key-Policy Attribute-Based Temporary Keyword Search Scheme" (KP-ABTKS), Lecture Notes in Networks and Systems, 2020.

[13] . Madhavi.K., G. Ramesh, G. Lavanya ”Load effectiveness on coverage-technique for test case prioritization in regression testing", International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-8 Issue-7 May, 2019.

[14] . Koshizuka, N.; Sakamura, K. Ubiquitous ID: Standards for ubiquitous computing and the Internet of Things.IEEE Pervasive Comput. 2010, 9, 98–101. Sensors 2018, 18, 2796 33 of 37

[15] . Want, R. An introduction to RFID technology. IEEE Pervasive Comput. 2006, 5, 25–33.

[16] . Wu, M.; Lu, T.J.; Ling, F.Y.; Sun, J.; Du, H.Y. Research on the architecture of Internet of Things. In Proceedings of the 3rd International Conference on Advanced Computer Theory and Engineering (ICACTE), Chengdu, China, 20– 22 August 2010; Volume 5, pp. V5-484–V5-487