# DDoS Attack Detection using Machine Learning Techniques in Cloud Computing Environments

**Marwane Zekri[1], Said El Kafhali[2], Noureddine Aboutabit[1] and Youssef Saadi[3]**

[1]IPOSI laboratory
National School of Applied Sciences, Hassan $1^{st}$ Univ, Settat, Morocco
*marwane.zekri@gmail.com, noure049@gmail.com*

[2]Computer, Networks, Mobility and Modeling laboratory
National School of Applied Sciences, Hassan $1^{st}$ Univ, Settat, Morocco
*said.elkafhali@uhp.ac.ma*

[3]Computer, Networks, Mobility and Modeling laboratory
Faculty of Sciences and Technology, Hassan $1^{st}$ Univ, Settat, Morocco
*youssadi@gmail.com*

***Abstract*—Cloud computing is a revolution in IT technology that provides scalable, virtualized on-demand resources to the end users with greater flexibility, less maintenance and reduced infrastructure cost. These resources are supervised by different management organizations and provided over Internet using known networking protocols, standards and formats. The underlying technologies and legacy protocols contain bugs and vulnerabilities that can open doors for intrusion by the attackers. Attacks as DDoS (Distributed Denial of Service) are ones of the most frequent that inflict serious damage and affect the cloud performance. In a DDoS attack, the attacker usually uses innocent compromised computers (called zombies) by taking advantages of known or unknown bugs and vulnerabilities to send a large number of packets from these already-captured zombies to a server. This may occupy a major portion of network bandwidth of the victim cloud infrastructures or consume much of the servers time. Thus, in this work, we designed a DDoS detection system based on the C.4.5 algorithm to mitigate the DDoS threat. This algorithm, coupled with signature detection techniques, generates a decision tree to perform automatic, effective detection of signatures attacks for DDoS flooding attacks. To validate our system, we selected other machine learning techniques and compared the obtained results.**

***Keywords—Security, Cloud Computing, DDoS attack, Vulnerability, Intrusion Detection, Machine Learning.***

## I. INTRODUCTION

Cloud computing refers to a type of Internet-based computing that provides shared pool of resources such as network bandwidth, memory, computer processing and user applications. Theses resources can be rapidly provisioned on demand over Internet [1]–[3] to End Users with less maintenance and less infrastructure cost. The cloud computing services can be categorized into three models [4]: Software-as-a-service (SaaS), Platform-as a Service (PaaS) and Infrastructure-as-a-Service (IaaS). Moreover, it can be deployed as private, public, community or hybrid cloud. Today, limiting the number of organizations willing to embrace the cloud wholeheartedly is one of the biggest challenges this technology is facing. DDoS is a type of aggressive attack that causes serious troubles on cloud servers.

According to [5], the term Denial of Service (DoS) was originally coined by Gligor in an operating system context [6], but since became widely adopted. A DoS attack involving more than one computer to target a victim in a coordinated manner is called a Distributed Denial of Service (DDoS) attack. This work focuses on using machine-learning techniques for detecting DDoS attacks.

The problem of attack detection using machine-learning techniques is not new to literature. While signature detection techniques can detect attacks based on signatures of already learnt attacks, anomaly detection techniques learn network traffic from a baseline profile and detect anomalies as ones that deviate significantly from the baseline profile. Signature detection techniques are effective against known attacks while anomaly detection has the ability to detect unknown and new attacks (zero-day). Data flow generated by attack presents irregular status. This makes DDoS attacks launched easily, prevented and tracked difficultly and so forth.

Furthermore, DDoS attacks have become one of the essential threats to network security. Following, we introduce machine learning in order to pinpoint our system and have some grasp of what machine learning is and how it is evolving. Machine Learning is a sub-set of artificial intelligence where computer algorithms can be used to autonomously learn from data. Machine learning severely affects most industries and the jobs within them. To arrive at this point advancement in the field of machine learning get passed by many major milestones [7]–[9].

In 1952 [10], Arthur Samuel wrote the first computer learn-

978-1-5386-1115-9/17/$31.00 ©2017 IEEE

ing program that is a checkers-playing program. Since then a lot of works were done. Frank Rosenblatt in [11], designed the first neural network in 1957. The nearest neighbor algorithm [12] was written in 1967 and so on until 1990s where works on machine learning shifted from a knowledge-driven approach to a data-driven approach. Scientists then begin creating programs for computers to analyze large amounts of data and draw conclusions. In 1997 [13], IBMs Deep Blue beats the world champion at chess. In 2006 [14], Geoffrey Hinton invented the term deep learning to explain new algorithms that let computers recognize and distinguish objects and text within images and videos. In 2011 [15], IBMs Watson beats its human competitors at Jeopardy. In 2014 [16], Facebook develops DeepFace, to recognize individuals from photos such as humans can do. In 2015 [17], Microsoft creates the Distributed Machine Learning Toolkit that enables the distribution of machine learning problems across multiple computers.

As the amount of data we produce continue to grow exponentially, the need for computers to analyze this large amount of data and to draw more efficient and accurate conclusions becomes a serious requirement. To this end, enhancing computers ability to process, analyze and learn from growing large amounts of data also requires more attention. As data grows and expands, decision tools need to evolve. In this context, we presented in [18] a novel mitigation system against the EDoS (Economic Denial of Sustainability) attacks. In other hands we focus in this paper on detection schemes we use machine learning techniques to detect and mitigate the DDoS attacks.

The main contributions of this paper can be summarized as follows:

- We proposes a faster and accurate DDoS detection attacks using a decision tree technique based on C4.5 algorithm.
- We Coupled our approach with signature detection techniques to provide high detection accuracy.
- We compared between other machine learning techniques to validate the obtained results.

The rest of paper is organized as follows. Section II summarizes previous studies related to DDoS attack detection using machine-learning techniques. In Section III we discuss the DDoS Attacks intend, launch methods and incentives, intrusion detection methodologies, and we present an overview on machine-learning techniques. In Section IV, we analyze the proposed DDoS detection architecture utilizing C.4.5 algorithm. Section V presents simulation environment, collected data and the obtained results. Finally, in Section VI, we conclude our works with future directions.

## II. Related Work

In this section, we will present a summary of existing literature on DDoS attack detection methods. We will briefly summarize the recent trends in DDoS attack detection mechanisms. Liao *et al.* [19] proposed a detection scheme based on Support Vector Machine (SVM). The detection is based on similarity of bots in accessing the web pages. They used feature like request frequency sequence to record the request patterns of users and apply rhythm-matching algorithm to identify similar patterns. Xiao *et al.* [20] proposed a detection scheme based on the property that the flows generated by the same software are likely to be correlated with each other. They used k-nearest neighbours algorithm to identify the flows that may have occurred from the same software or bots. However, if an attacker uses different configuration parameters for initiating an attack from the bots then they may generate non-similar flows.

In [21], authors propose a new anomaly-based DDoS detection method based on various features of packets attack by analyzing them using a particular type of neural networks called the Radial Basis Function (RBF) neural networks. The method can be applied at the edge router of victim networks. They used seven feature vector to activate RBF neural network at each time-window. The RBF neural network is applied to classify data into two categories: normal and attack. If the incoming traffic is recognized as attack traffic, the source IP addresses of the attacked packets are sent to the Filtering and to the Attack Alarm Modules for further actions. Otherwise, if the traffic is recognized as normal, it is sent to the destination. A DDoS attack detection model based on data-mining algorithm is presented in [22]. Authors used FCM cluster algorithm and a priori association algorithm to extract network traffic model and network packet protocol status model and set the threshold for detection model. The authors in [23], proposed the concept of DDoS attack detection using decision tree and grey relational analysis. The detection of attack from normal situation is perceived as a classification problem where they proposed 15 different attributes that not only monitor the incoming/outgoing packet/bytes rate but also compile the TCP SYN and ACK flag rate to describe the traffic flow pattern. The decision tree technique was applied taking these attributes as tests to detect abnormal traffic flow.

Jie-Hao *et al.* [24] used Artificial Neural Networks(ANN) to detect DDoS attacks where they compared the detection outcome with decision tree, ANN, entropy and Bayesian. The authors identified users requests to a specific resource and their communicative data. Then samples of such requests are sent to the detection systems to be judged for abnormalities. In addition, Liu *et al.* [25] have used another type of neural networks named Learning Vector Quantization (LVQ) neural networks to detect attacks. This is a supervised version of quantization which can be used for pattern recognition, multiclass classification and data compression tasks. The datasets used in the experiments were converted into numerical form and then given as inputs to the neural network. The authors in [26], have introduced a Probabilistic Neural Network Based Attack Traffic Classification to detect different DDoS attacks. However, the authors focused on separating Flash Crowd Event from Denial of Service Attacks. As part of their research, they have used Bayes decision rule for Bayes inferences coupled with Radial Basis Function Neural Network (RBFNN) for classifying DDoS attack traffic and normal traffic.

Detection accuracy can be improved by combining SVM with other techniques. Li *et al.* [27], designed an intelligent module for network intrusion prevention system with a combination of SNORT and configurable firewall. The SVM classifier is also used with SNORT to reduce false alarm rate and improve accuracy of Intrusion Prevention System (IPS). Distributed Time Delay Neural Network (DTDNN) [28] has higher detection accuracy for most of the network attacks. DTDNN is a simple and efficient solution for classifying data with high speed and fast conversion rates. Accuracy of this approach can be improved by combining it with other soft computing techniques. The authors in [29], proposed a DDoS attack detection model based on protocol analysis and cluster. This model uses data mining algorithm to analyze the protocol information elements piggybacked in packets. The advantages is that the attack detection method does not needs any manual construction of data while keeping a comparatively high detection rate. However, the number of network connections in unit time will range from ten thousand to one million for a large scale network.

## III. DDoS attacks and detection Methodologies

Despite the fast increasing popularity of cloud services, ensuring the security and availability of data, resources and services remains an ongoing research challenge. Distributed denial of service (DDoS) attacks are not a new threat, It is major security issue and a wide topic of ongoing research interest. In this section we discuss the various DDoS intend and Launch methods that could be used to conduct or facilitate DDoS attacks, as well as reviewing intrusion Detection Methodologies and defense strategies.

### *A. DDoS Attack*

*1) DDoS intend and Launch methods:* DoS attacks are intended attempts to stop legitimate users from accessing a specific network resource.The Open Systems Interconnection Model (OSI model), is useful in understanding the types of DDoS attacks we are dealing with .DDoS attacks target specific layers of a network connection(application layer attacks target layer 7, protocol layer attacks target layers 3 and 4). the first DDoS attack incident was reported [30].

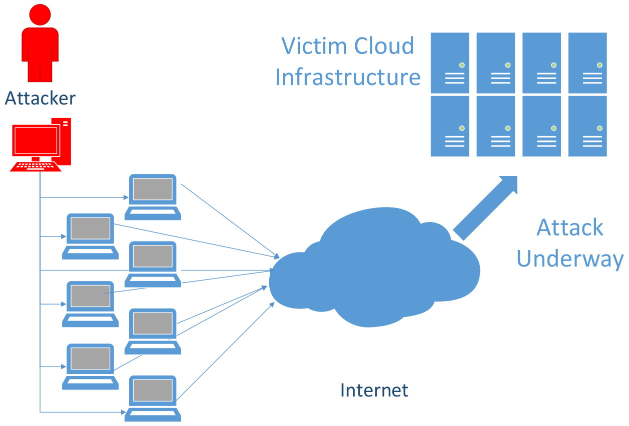

Fig. 1. Typical Architecture of DDoS Attacks

Currently, there are two main methods to trigger a DDoS attacks in the Internet. The first is to send some malformed packets to the victim (i.e., vulnerability attack). The second method, involves an attacker trying to do one or both of the following:

- Disrupt a legitimate users connectivity by exhausting bandwidth, router processing capacity or network resources. These are essentially network/transport-level flooding attacks [31]. (i.e., flooding attacks)
- Disrupt legitimate users services by exhausting the server resources (e.g., sockets, CPU, memory, disk/database bandwidth, and I/O bandwidth) These essentially include application-level flooding attacks.

*2) The attacker's incentives:* DDoS attackers are usually motivated by various justifications. Analyzing the attackers incentives help to stop and respond to these attacks [32].

- **Economical/Financial gain:** A major concern of corporations generally performed by frustrated individuals, possibly with lower technical skills.
- **Intellectual Challenge:** The attacks are usually young hacking enthusiasts who want to show off their capabilities to experiment and learn how to launch various attacks.
- **Cyber warfare:** This category of Attackers are usually politically motivated to attack a wide range of critical sections of another country.
- **Ideological belief:** Attackers who belong to this category are motivated by their ideological beliefs to attack their targets, political incentives have led to recent sabotages, like the year 2007 sabotage attack in Estonia [33], Iran 2009 [34], WikiLeaks 2010 [35].

### *B. Intrusion Detection Methodologies*

Here we introduce the intrusion detection methodologies [36]–[38]. These methodologies are classified into three major categories: Signature-based Detection (SD), Anomaly-based Detection (AD) and Stateful Protocol Analysis (SPA). Following, we present the pros and cons of the three detection methodologies.

**Signature-based Detection:** The Signature-based ID systems detects intrusions by observing events and identifying patterns which match the signatures of known attacks. An attack signature defines the essential events required to perform the attacks and the order in which they must be performed. Moreover, it detects only attacks whose signatures are previously stored in database. For efficiency purpose, the signatures must be updated regularly. Just as new threats are released regularly, creating the need for signature updates, new threats against your hosts are discovered regularly. This method work well against only the fixed behavioral pattern. They fail to deal with attacks created by human or a worm with self-modifying behavioral characteristics.

**Anomaly-based Detection:** The AD has attracted many researchers due to its capability of detecting novel attack. The detecting is based on defining the network behavior. The network behavior is in Conformity with the predefined behavior. Then it is accepted or else it triggers the event in the anomaly detection. The accepted network behavior can be prepared or learned by the specifications of the network administrators. The major advantage of AD compared to signature-based engines is that a novel attack for which a signature does not exist can be detected if it falls out of the normal traffic patterns.

**Stateful Protocol Analysis:** The SPA indicates that IDS could know and trace the protocol states (e.g., pairing the requests with replies). AD adopts pre-loaded network or host specific profiles, whereas SPA depends on vendor developed generic profiles to specific protocols that specify how particular protocols should and should not be used. SPA provides important capabilities for understanding and responding to attacks. The SPA as an intrusion detection method relies upon well-defined and well-behaved protocol models. In cases where a protocol is proprietary, poorly defined, or a vendor implementation deviates from the standard, SPA becomes less accurate.

### *C. Machine Learning Techniques*

In this section, we briefly describe the various machine learning algorithms and the problem domains they are frequently used in. Many decision tree and rule induction algorithms have already been suggested in the literature. The Naive Bayes algorithm [39] is a probabilistic classifier, it assumes that the effect of a variable values on a given class is independent of the values of other variables. This assumption is called class conditional independence. Decision tree is one of the most well-known and used classification algorithms. C4.5 algorithm [40] which was developed by Ross Quinlan is the most popular tree classifier. This algorithm is based on ID3 (Iterative Dichotomiser 3) algorithm that tries to find a small decision tree. The decision tree generated by C4.5 can be used for classification, and it often referred to as a statistical classifier. Authors of the Weka machine learning software [41], described the C4.5 algorithm as a landmark decision tree program that is probably the machine learning workhorse most widely used in practice to date [42]. K-Mean Clustering [43] assignment of the data points to clusters depends upon the distance between cluster centroid and data point.

The k-NN (k-Nearest Neighbors) algorithm [44] is a similarity-based learning algorithm and is known to be highly effective in various problem domains, including classification problems. In classification and regression, SVM (Support Vector Machines) [45] is the most popular method for machine learning tasks. SVM can be applied not only to classification problems but also to the case of regression problems. FCM Clustering (Fuzzy C-Means clustering) [46] is a method of clustering which allows one piece of data to belong to two or more clusters. This method is frequently used in pattern recognition. The Neural Networks (NNs) [47] are mathematical representations inspired by the functioning of the human brain. Many applications of NNs have been suggested in the literature [48]: Character Recognition, Image Compression, Stock Market Prediction, Medicine, Electronic Nose, Security, and Loan Applications and many other examples.

Machine learning techniques are frequently used for anomaly detection [49]. They have received considerable attention among the intrusion detection researchers to address the weaknesses of knowledge base detection techniques. Experiment done by [50] show that C4.5 is more stable than k-NN. Another experiment on developed three intrusion detection models based Multi-Layer Perceptron (MLP), C4.5, and SVM classifiers [51] showed that C4.5 is the best method in terms of detection accuracy and minimum training time; it achieved the accuracy rate of (99.05%). For this reason, we choose the C4.5 algorithm to detect the DDoS attacks in our proposed model.

## IV. Proposed Model

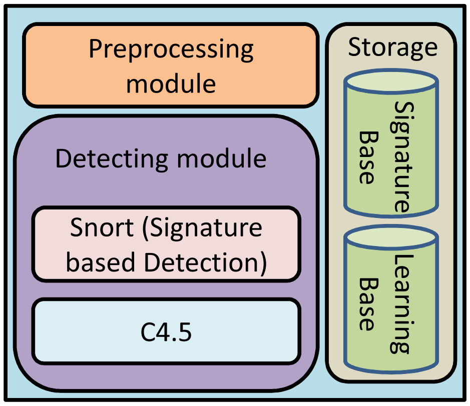

Fig. 2. Proposed Model

As shown in Figure 3, Pre-processing module processes captured packets in a specific format by removing redundant information that has very low correlation with detection. Such as signature based Detection, it matches a given network event with the rules stored in knowledge base and detects known attack efficiently. One advantage of using this technique is that we can easily update knowledge base without modifying the existing rules. It involves data collection (stored in learning base) related to the behavior of legitimate user within a period of time, and then applies machine learning algorithm to the this data, which determines whether that user is legitimate or not.

### *A. Model Goals*

In this subsection, we describe the goals that we aim to achieve by applying the proposed model. We have the following objectives:

- Low computational cost,
- Faster detection rate,
- Detection of network DDoS in Cloud environment,
- Scalability,
- Low false positives and false negatives,
- High accuracy.

### B. Model Theoretical Background

To compare our results we use Naive Bayes classifier [52] for anomaly detection, whereas Snort [53] is used as signature based detection. Snort is an open source IDS which uses signature based technique for detecting attacks. It is widely used and beside it can run on multiple platforms (i.e . GN U/Linux, Windows). Moreover, it is constantly updated. It captures network data packets and checks their content with the predefined known attack patterns for any correlation. Snort is most used to prevent system from known attacks. It is a statistical classifier that predicts the probability of a given network events belong to particular class (normal or abnormal). It has higher accuracy and speed than other classifiers (e.g. decision tree classifier and neural network classifier). The $X$ is a given packet. H is hypothesis, such that $X$ belongs to class $C$. We need to determine the probability $P\left(H|X\right)$ that the hypothesis holds the packet $X$. $P\left(X\right)$ is the initial probability of H. $P\left(X\right)$ is the probability that packet is observed. $P\left(X|H\right)$, the probability of observing packet $X$, given that the hypothesis holds. Using Bayes theorem, the probability $P\left(H|X\right)$ of a hypothesis $H$, on given packet $X$ can be derived by equation (1) as follows

$$P\left(H|X\right) = \frac{P\left(X|H\right)P\left(H\right)}{P\left(X\right)} \tag{1}$$

- **Protocol:** is a set of rules that govern and describes how data packets move through a network.
- **Flag:** can help to troubleshoot a connection.
- **Service:** is a connection type transmitting messages/files.
- **Land:** if the source and destination IP are same, then land variable is set to 1.
- **TTL:** initial Time To Live (TTL) values, 30, 32, 60, 64, 128. This set of initial TTL values covers most of the popular OSes. we can determine the initial TTL value of a packet by selecting the smallest initial value in the set that is larger than its final TTL. For example, if the final TTL value is 112, the initial TTL value is 128.

TABLE I
SAMPLE TRAINING DATASET

| Land | Service | Protocol | Flag | TTL | Class |
|---|---|---|---|---|---|
| 0 | http | TCP | SYN | TTL≥128 | Normal |
| 0 | http | TCP | SYN | TTL≤128 | Attack |
| 1 | ftp | UDP | | TTL≥128 | Attack |
| 0 | smtp | ICMP | | TTL≥128 | Normal |
| 0 | http | UDP | | TTL≥128 | Normal |
| 1 | http | TCP | SYN | TTL≥128 | Attack |
| 0 | http | UDP | | TTL≥128 | Normal |
| 1 | ftp | ICMP | | TTL≥128 | Normal |
| 1 | ftp | ICMP | | TTL≥128 | Attack |
| 0 | ftp | ICMP | | TTL≤128 | Attack |

We adopt the C4.5 algorithm to construct the decision tree. C4.5 chooses the attribute as the splitting criterion according to the entropy-based gain ratio in order to overcome the overfitting problem. First, C4.5 defines $Info(D)$, that represents the entropy of the training data set $D$ and the probability that one random instance from $D$ belongs to a class $C_j$ (there are four classes in our system: Normal, TCP SYN attack, UDP attack, and ICMP attack and one traffic signature aggregated per 1 minute would be considered as one instance in our system). We can define $Info(D)$ in equation (2) as follows

$$Info\left(D\right) = -\sum_{j=1}^{k}\left[\frac{|D_j|}{|D|}\log_2\left[\frac{|D_j|}{|D|}\right]\right] \tag{2}$$

The gain information for an attribute, $Gain(X)$, measures the quantity of information that is gained by partitioning $D$ according to the attribute $X$ (we treat the format of traffic signature defined as the attributes in our system). The $Gain(X)$ is expressed in equation (3) as

$$Gain\left(X\right) = Info\left(D\right) - \sum_{j=1}^{k}\left[\frac{|D_j|}{|D|}Info\left(D\right)\right] \tag{3}$$

where $D_i$ represents the number of instances in the specific attribute. We define the gain ratio in equation (4) as follows

$$Gain_{ratio}\left(D\right) = \frac{Info\left(D\right)}{-\sum_{j=1}^{k}\left[\frac{|D_j|}{|D|}\log_2\left[\frac{|D_j|}{|D|}\right]\right]} \tag{4}$$

The attribute with the largest gain ratio is selected as the splitting criterion in the decision tree. Based on the selected attribute, the training data set is then divided into several subsets. Another attribute is similarly selected and each subset is further split. The splitting procedure is repeated until all the data in a subset belong to the same class or the gain ratios of all the attributes are the same.

The construction procedure can be summarized as the following:

1) The first step is to select the attribute with has the largest gain ratio as the splitting criterion, and create a branch for each possible value of the selected attribute.
2) Divide the instances in the training data set into subsets according to the selected attribute.
3) Repeat steps 1 and 2 for each branch. In our implementation, we define four classes: normal, TCP SYN flooding, UDP flooding, ICMP flooding and attributes derived from the traffic signatures. A decision tree is then constructed from the training data set. According to the decision tree, the incoming traffic is classified.

## V. SIMULATION RESULTS AND DISCUSSION

### A. Experiment environment

We introduce in this section the important concepts, principles and terminology relevant for experiment design. Subsequently, we explain the main elements of our approach. No public cloud is used in our simulation. Any DDoS attack can create an impact on the public cloud, so it is not possible to show this simulation in large cloud networks. Furthermore, the public cloud (Azure, AWS, Google Cloud Platform etc) have their own defense mechanisms. For the lack of resources, and both ethical and legal issues the simulation was done in a virtual environment with VMs and virtual LAN. The experimental setup is OpenStack juno an open source software for building public, private, and hybrid clouds and capable of many functions associated with cloud computing such as flexible provisioning of VMs, cloning and snapshotting VM images, and online/offline migration We also used Oracle Virtual Box as the base environment for all the system, Hping3 used to generate attacks. For monitoring, we used (Wireshark, IP Traffic monitoring, etc). Our testbed is composed with the target PMs (Physical Machines) connected to a router through an internal IP router and the router is connected to the Internet through an external IP router. The type of attacks used in this experiment to obtain detection result is DDoS flooding attacks. For the generation of normal network traffic, we used an parameterized python-scripts the Classification of attack and normal traffic done using the proposed method.

TABLE II
DDoS ATTACK DETECTION RESULT IN DIFFERENT DURATION

| Method used | Correct Classification (%) | Detection Time (S) |
|---|---|---|
| Naive Bayesian | 91.4 | 1.25 |
| C4.5 | 98.8 | 0.58 |
| K-Means | 95.9 | 1.12 |

Table II shows the correct classification and the attack detection time, in term of the running time 0.58s for C4.5 and 1.1s is time taking by Naive Bayesian. For correct classification C4.5 work better than Naive Bayesian, it has a score of 98.8 (%) in the other hand Naive Bayesian scored 91.4 (%). The obtained results approve that the C4.5 is the better method regarding the choice of classification technique used.

TABLE III
F-MEASURE DETAILS OF CLASSIFIERS

| | TP | FP | TN | FN | F-Measure |
|---|---|---|---|---|---|
| Naive Bayesian | 282 | 28 | 253 | 25 | 0.914 |
| C4.5 | 298 | 2 | 270 | 5 | 0.988 |
| K-Means | 285 | 24 | 264 | 0 | 0.959 |

F-measure is calculated based on the precision and recall. The calculation is as follows:

$$Precision = \frac{TP}{TP + FP} \tag{5}$$

$$Recall = \frac{TP}{TP + FN} \tag{6}$$

$$F - Measure = \frac{2 * Precision * Recall}{Precision + Recall} \tag{7}$$

Where TP represent the number of true positives, FP the number of false positives and FN the number of false negatives. Precision is determined by defined as the fraction of elements correctly classified as positive out of all the elements the algorithm classified as positive, whereas recall is defined as the fraction of elements correctly classified as positive out of all the positive elements. From the analysis of DDoS attacks in this experiment, it is found that this system has a high detection and efficiency rate, the detection rate of it could reach more than 98%. Moreover, with the increase in the duration of DDoS attacks, the higher is the attack detection rate of this system. The function test result of this system shows that it could meet the daily detection needs well.

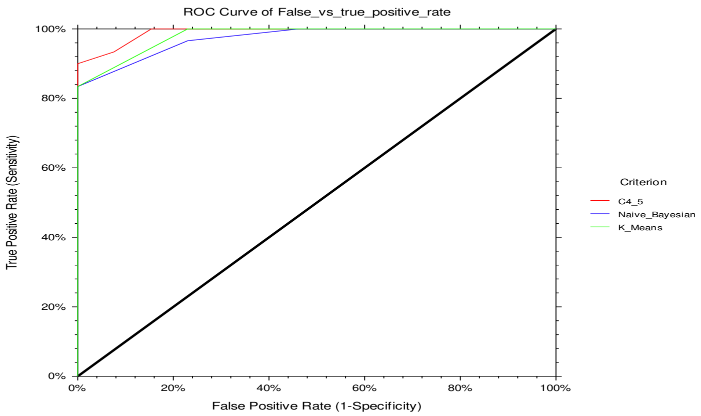

Fig. 3. False vs true positive rate

## VI. Conclusion and Future work

In this work, we presented a DDoS detection system design based on machine learning and signature detection techniques. Attack detecting is a wide topic of research for cloud computing that aims to make the Cloud a secure and trusted platform for the delivery of future Internet of Things. In this paper, we focused on the flooding-based attack targeting layer 3 and layer 4 in the OSI 7-layer model. We applied the decision tree (C4.5) technique to detect abnormal traffic flow. A comparative analysis of various machine learning strategies and algorithms was presented. From the obtained results, we concluded that our proposed approach, in which C4.5 algorithm is adopted to detect DDoS attacks, gives more accurate results in comparison with others machine learning algorithms. Finally, as a future work, we are planning to develop a prototype system based on the proposed mechanism for real-time attack traffic detection and mitigation aimed to security challenges.

## References

[1] A. Shawish and M. Salama, “Cloud computing: paradigms and technologies,” in *Inter-cooperative collective intelligence: Techniques and applications*. Springer, 2014, pp. 39–67.

[2] S. El Kafhali and K. Salah, “Stochastic modelling and analysis of cloud computing data center,” in *20th Conference on Innovations in Clouds, Internet and Networks (ICIN)*. IEEE, 2017, pp. 122–126.

[3] P. Mell, T. Grance *et al.*, “The nist definition of cloud computing,” 2011.

[4] S. El Kafhali and K. Salah, “Performance analysis of multi-core vms hosting cloud saas applications,” *Computer Standards & Interfaces*, 2017.

[5] K. Lakshminarayanan, D. Adkins, A. Perrig, and I. Stoica, "Taming ip packet flooding attacks," *ACM SIGCOMM Computer Communication Review*, vol. 34, no. 1, pp. 45–50, 2004.

[6] V. D. Gligor, "A note on denial-of-service in operating systems," *IEEE Transactions on Software Engineering*, no. 3, pp. 320–324, 1984.

[7] S. C. Hoi, J. Wang, and P. Zhao, "Libol: A library for online learning algorithms," *The Journal of Machine Learning Research*, vol. 15, no. 1, pp. 495–499, 2014.

[8] K. R. Foster, R. Koprowski, and J. D. Skufca, "Machine learning, medical diagnosis, and biomedical engineering research-commentary," *Biomedical engineering online*, vol. 13, no. 1, p. 94, 2014.

[9] B. Marr, "A short history of machine learning – every manager should read," *https://www.forbes.com/sites/bernardmarr/2016/02/19/a-short-history-of-machine-learning-every-manager-should-read/3109523415e7*, 2016.

[10] K. Solanki, A. Dhankar *et al.*, "A review on machine learning techniques." *International Journal of Advanced Research in Computer Science*, vol. 8, no. 3, 2017.

[11] I. Basheer and M. Hajmeer, "Artificial neural networks: fundamentals, computing, design, and application," *Journal of microbiological methods*, vol. 43, no. 1, pp. 3–31, 2000.

[12] J. Derrac, S. García, and F. Herrera, "Fuzzy nearest neighbor algorithms: Taxonomy, experimental analysis and prospects," *Information Sciences*, vol. 260, pp. 98–119, 2014.

[13] M. Campbell, A. J. Hoane, and F.-h. Hsu, "Deep blue," *Artificial intelligence*, vol. 134, no. 1-2, pp. 57–83, 2002.

[14] A. Lee, "The meaning of alphago, the ai program that beat a go champ," *http://www.macleans.ca/society/science/the-meaning-of-alphago-the-ai-program-that-beat-a-go-champ/*, 2016.

[15] D. A. Ferrucci, "Introduction to this is watson," *IBM Journal of Research and Development*, vol. 56, no. 3.4, pp. 1–1, 2012.

[16] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, "Deepface: Closing the gap to human-level performance in face verification," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014, pp. 1701–1708.

[17] Y.-h. Tian, X.-l. Chen, H.-k. Xiong, H.-l. Li, L.-r. Dai, J. Chen, J.-l. Xing, X.-h. Wu, W.-m. Hu, Y. Hu *et al.*, "Towards human-like and transhuman perception in ai 2.0: a review," *Frontiers of Information Technology & Electronic Engineering*, vol. 18, no. 1, pp. 58–67, 2017.

[18] M. Zekri, S. El Kafhali, M. Hanini, and N. Aboutabit, "Mitigating economic denial of sustainability attacks to secure cloud computing environments," *Transactions on Machine Learning and Artificial Intelligence*, vol. 5, no. 4, pp. 473–481, 2017.

[19] Q. Liao, H. Li, S. Kang, and C. Liu, "Application layer ddos attack detection using cluster with label based on sparse vector decomposition and rhythm matching," *Security and Communication Networks*, vol. 8, no. 17, pp. 3111–3120, 2015.

[20] P. Xiao, W. Qu, H. Qi, and Z. Li, "Detecting ddos attacks against data center with correlation analysis," *Computer Communications*, vol. 67, pp. 66–74, 2015.

[21] R. Karimazad and A. Faraahi, "An anomaly-based method for ddos attacks detection using rbf neural networks," in *Proceedings of the International Conference on Network and Electronics Engineering*, 2011, pp. 16–18.

[22] R. Zhong and G. Yue, "Ddos detection system based on data mining," in *Proceedings of the 2nd International Symposium on Networking and Network Security, Jinggangshan, China*, 2010, pp. 2–4.

[23] Y.-C. Wu, H.-R. Tseng, W. Yang, and R.-H. Jan, "Ddos detection and traceback with decision tree and grey relational analysis," *International Journal of Ad Hoc and Ubiquitous Computing*, vol. 7, no. 2, pp. 121–136, 2011.

[24] J. Li, Y. Liu, and L. Gu, "Ddos attack detection based on neural network," in *2nd International Symposium on Aware Computing (ISAC),*. IEEE, 2010, pp. 196–199.

[25] V. Akilandeswari and S. M. Shalinie, "Probabilistic neural network based attack traffic classification," in *Fourth International Conference on Advanced Computing (ICoAC)*. IEEE, 2012, pp. 1–8.

[26] J.-H. Chen, M. Zhong, F.-J. Chen, and A.-D. Zhang, "Ddos defense system with turing test and neural network," in *IEEE International Conference on Granular Computing (GrC)*. IEEE, 2012, pp. 38–43.

[27] H. Li and D. Liu, "Research on intelligent intrusion prevention system based on snort," in *International Conference on Computer, Mechatronics, Control and Electronic Engineering (CMCE)*, vol. 1. IEEE, 2010, pp. 251–253.

[28] L. M. Ibrahim, "Anomaly network intrusion detection system based on distributed time-delay neural network (dtdnn)," *Journal of Engineering Science and Technology*, vol. 5, no. 4, pp. 457–471, 2010.

[29] N. Gao, D.-G. Feng, and J. Xiang, "A data-mining based dos detection technique." *Jisuanji Xuebao(Chinese Journal of Computers)*, vol. 29, no. 6, pp. 944–951, 2006.

[30] P. J. Criscuolo, "Distributed denial of service: Trin00, tribe flood network, tribe flood network 2000, and stacheldraht ciac-2319," CALIFORNIA UNIV LIVERMORE RADIATION LAB, Tech. Rep., 2000.

[31] S. T. Zargar, J. Joshi, and D. Tipper, "A survey of defense mechanisms against distributed denial of service (ddos) flooding attacks," *IEEE communications surveys & tutorials*, vol. 15, no. 4, pp. 2046–2069, 2013.

[32] N. Fultz and J. Grossklags, "Blue versus red: Towards a model of distributed security attacks." in *Financial Cryptography*, vol. 5628. Springer, 2009, pp. 167–183.

[33] B. Zdrnja, "Slowloris and iranian ddos attacks," *http://isc.sans.edu/diary.html?storyid=6622*, 2009.

[34] E. Schonfeld, "Wikileaks reports it is under a denial of service attack," *http://techcrunch.com/2010/11/28/wikileaks-ddos-attack*, 2010.

[35] R. SATTER, "Us general: We hacked the enemy in afghanistan," *http://csulauniversitytimes.com/1559/news/us-general-we-hacked-the-enemy-in-afghanistan/*, 2012.

[36] H.-J. Liao, C.-H. R. Lin, Y.-C. Lin, and K.-Y. Tung, "Intrusion detection system: A comprehensive review," *Journal of Network and Computer Applications*, vol. 36, no. 1, pp. 16–24, 2013.

[37] S. Axelsson, "Intrusion detection systems: A survey and taxonomy," Technical report, Tech. Rep., 2000.

[38] H. Debar, M. Dacier, and A. Wespi, "A revised taxonomy for intrusion-detection systems," *Annals of Telecommunications*, vol. 55, no. 7, pp. 361–378, 2000.

[39] A. L. Blum and P. Langley, "Selection of relevant features and examples in machine learning," *Artificial intelligence*, vol. 97, no. 1, pp. 245–271, 1997.

[40] J. R. Quinlan, *C4. 5: programs for machine learning*. Elsevier, 2014.

[41] G. Holmes, A. Donkin, and I. H. Witten, "Weka: A machine learning workbench," in *Proceedings of the 1994 Second Australian and New Zealand Conference on Intelligent Information Systems*. IEEE, 1994, pp. 357–361.

[42] I. H. Witten, E. Frank, M. A. Hall, and C. J. Pal, *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann, 2016.

[43] J. A. Hartigan and M. A. Wong, "Algorithm as 136: A k-means clustering algorithm," *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, vol. 28, no. 1, pp. 100–108, 1979.

[44] J. M. Keller, M. R. Gray, and J. A. Givens, "A fuzzy k-nearest neighbor algorithm," *IEEE transactions on systems, man, and cybernetics*, no. 4, pp. 580–585, 1985.

[45] M. A. Hearst, S. T. Dumais, E. Osuna, J. Platt, and B. Scholkopf, "Support vector machines," *IEEE Intelligent Systems and their applications*, vol. 13, no. 4, pp. 18–28, 1998.

[46] J. C. Bezdek, R. Ehrlich, and W. Full, "Fcm: The fuzzy c-means clustering algorithm," *Computers & Geosciences*, vol. 10, no. 2-3, pp. 191–203, 1984.

[47] S. S. Haykin, S. S. Haykin, S. S. Haykin, and S. S. Haykin, *Neural networks and learning machines*. Pearson Upper Saddle River, NJ, USA:, 2009, vol. 3.

[48] E. Roberts, "Applications of neural networks," *https://cs.stanford.edu/people/eroberts/courses/soco/projects/neural-networks/Applications/index.html*, 2000.

[49] B. Lane, M. Poole, M. Camp, and J. Murray-Krezan, "Using machine learning for advanced anomaly detection and classification," in *Advanced Maui Optical and Space Surveillance Technologies Conference*, 2016.

[50] H. Ismanto and R. Wardoyo, "Comparison of running time between c4. 5 and k-nearest neighbor (k-nn) algorithm on deciding mainstay area clustering," *International Journal of Advances in Intelligent Informatics*, vol. 2, no. 1, pp. 1–6, 2016.

[51] A. F. Sheta and A. Alamleh, "A professional comparison of c4. 5, mlp, svm for network intrusion detection based feature analysis," in *The International Congress for global Science and Technology*, vol. 47, 2015, p. 15.

[52] W. I. D. Mining, "Data mining: Concepts and techniques," *Morgan Kaufinann*, 2006.

[53] M. Roesch *et al.*, "Snort: Lightweight intrusion detection for networks." in *Lisa*, vol. 99, no. 1, 1999, pp. 229–238.