Detection of DDoS Attacks Using Machine Learning in Cloud Computing

Vishal Sharma(✉), Vinay Verma, and Anand Sharma

Department of Computer Science and Engineering, Mody University of Science and Technology, Lakshmangarh, Sikar 332311, India
er.vishul983@gmail.com, ervinayv@gmail.com, anand_glee@yahoo.co.in

Abstract. Cloud Computing is basically the use of software and hardware to provide a service over an internet network. Users use applications or can access files from any device with the help of cloud computing. The main thing is that device must be connected through the Internet. Cloud computing has many advantages like scalability, less maintenance, virtualization and requested resources to the users with reduced infrastructure cost, and greater flexibility. It faces many drawbacks like security attack as Distributed Denial of Service (DDoS).

DDoS attack is well-defined as a way of attack that includes multiple conceded computer systems attack a goal, like a server, any resource and website, and due to this a denial of service for the end users of the intended resource. The fake connection requests, flooding of inward messages, or distorted packets forces the whole system to slow down and shut down, in that way denying service to genuine end users and systems. In this paper we have analyzed and proposed the machine learning algorithms for detecting DDoS attack in cloud computing environment. This paper is using isolation forest anomaly detection technique and then the correlation will be used to for detection of DDoS attack.

Keywords: DDOS attack · Machine learning · Cloud computing

1 Introduction

Cloud computing defined as an Internet-based computing that gives ability to share a wide group of resources like memory, network bandwidth, user applications, and processing of computer with less infrastructure cost and less maintenance [1–3]. The services of the cloud computing can be characterized into the three models [4]: Platform-as a Service (PaaS), Software-as-a-service (SaaS), and Infrastructure-as-a- Service (IaaS).

It is arranged in any network as public based, private based, and community based or hybrid types of cloud. The cloud service providers are hosted the services for the businesses or the ultimate users to exploit the uses over a connection of network at the data centre. They are the companies that are offering distinct services in the cloud environment. In the area of cloud computing the security attacks is one of the disadvantages. This problem is because of the storage of data at diverse geographical extents in the cloud computing environment.

A. K. Luhach et al. (Eds.): ICAICR 2019, CCIS 1076, pp. 260–273, 2019.
https://doi.org/10.1007/978-981-15-0111-1_24

2 Attacks on Cloud

There are many security attacks in the area of cloud. There are numerous attacks is happened in the cloud like attack of Denial of service, attack of Malware insertion, attack of side channel, attack of Man in the middle and the attack of the type of authentication those we'll conversed underneath. There are various portions of the area of cloud as the data storage, during the transaction of the data, during the utilization of resources and resource sharing on which the attack might be happened. Because of enormous increase in the use of the cloud facilities that's why this is also may be one reason for the attack.

2.1 Denial of Service (DoS)

In this type of attack the massive amount of the service requests from the attacker is overloaded to the targeted cloud system for that reason that halts it from responding to the forthcoming new requests to its ultimate users. Bestowing to the several security association of cloud, this type of cloud is very much defenseless to this type of attack. This attack type can be classified into two parts as the DoS attack and other one is the DDoS (Distributed denial of service attack). When any particular system and any particular network are doing this type of attack it is known as the DoS attack and when it is done by numerous systems and the numerous networks it is known as the Distributed denial of service attack (DDoS). The DDoS attack is further categorized as based attack on volume, attacks of protocol, and attack of Application layer.

2.2 Injection of Malware

In this type of attack in targeted cloud computing system the attacker tries to insert the malicious type of service or the malicious virtual machine. Then the cloud computing system must act so as to trusts that it is a usable service which is created by the attacker. Then the cloud server redirects automatically entirely demands to this malicious service if the attacker succeeds to do this. Now the service requests of the victim services can be accessed by the attacker.

2.3 Side Channel Attack

In this type of attack the attacker after efforts to adjustment the complete system by injecting a malicious virtual machine close to the goal system dispatches the side channel types of attack. By This type of attacks the attacker tries to retrieve private data without non-exhaustive manner and some specific access. Due to this reason it gives larger effect than any other types of attacks.

2.4 Authentication Attack

In this type of attack the authentication portion is the main focuses of the cloud services. Primary authentication in most of the services is username and the password. It is a kind of the knowledge-based authentication. Secret questions Sharing, keys of

site and virtual keyboards is called as secondary authentication which is used by secure operational organizations like as the financial enterprise. Further we classified the attacks of authentication which is discussed below.

2.4.1 Brute Force Attack

In this type of attack we have to use a hit and trial method; to crack the actual password we have to test all possible types of combinations of the password.

2.4.2 Key Loggers

In this type of attack the attacker uses a type of software suite, which is track the activities of the end user by records each one single key which is pressed.

2.4.3 Phishing Attack

In view of this type of the attack the attacker to acquire the passwords and the PIN of the end user, alter the end user to the false websites; Phishing attack is a web-based type of attack.

2.5 Man-in-the-Middle-Attack

In this type of attack interrupts the message in the key exchange by replacing its own key for the wished one by the attacker, however the both actual end users are quiet communicating usually. The source doesn't identify the attacker which has received the message directed by him and he can approach the sender's data and the attacker can alter this message beforehand this data to the receiver.

3 DDoS in Cloud Computing

The concept of Denial of Service was at first conceptualized by Gligor in an operating system environment [5, 6], but subsequently it widely accepted. The Denial of Service stops the genuine users to access the resources in the specific network. The DoS attack is generally categorized in two parts first is Network level and other one is Application level. In the Network level attacks the attacker generally deactivates the genuine users' connectivity by draining resources of the network. In the Application level DoS attacks the attacker deactivates the services by draining the server resources.

A few DoS prevention methods feel necessity for the client to solve a particular challenge earlier just as proof-of-work. An advanced version of DoS attacks is Distributed Denial of Service attacks which are launched by several sources targeting the same victim. To launch DDoS attacks, Attackers entered into the target machines, take control of these machines and use machines as secondary victim to attack the primary victim. The attacker first uses some scanning techniques to compromise a network of vulnerable nodes called a botnet. Then the attacker sends the DDoS attack command to a botnet and forces it to launch the attack [7]. With this type of physical bots,

Distributed Denial of Service attacks are also exhausted on service clouds by renting many virtual machines or computers and using them as virtual machine bots to attack the external world [8]. In short, DDoS attacks are very easy to launch but extremely hard to trace back to the actual attackers [9]. Figure 1 illustrates how DDoS attacks are launched.

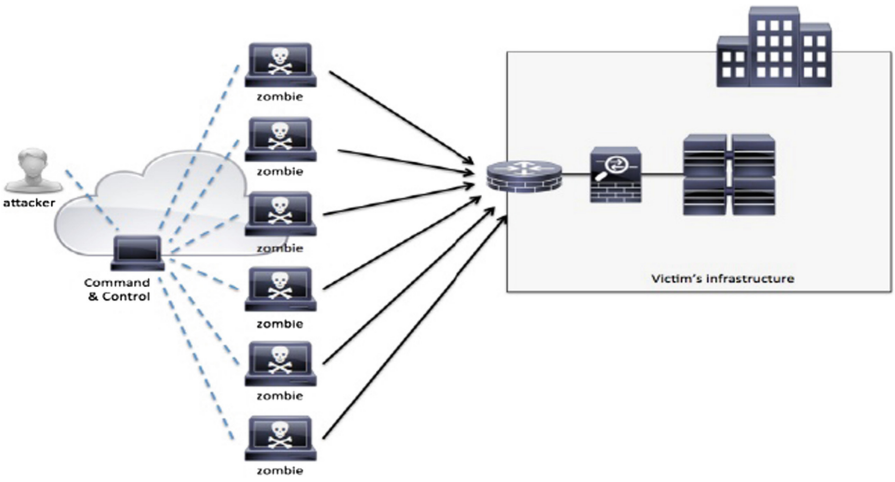

Fig. 1. DDoS attack

These attacks drastically disgrace the prestige of the cloud service provider. Due to this the cloud service providers go for large financial losses. At east coast USA scheduled 21st Oct 2016, DDoS attack down the Internet connection.

Prior to that in mid-September, 2016, the biggest and most practical DDoS attacks affected the JP Morgan Chase, Bank of America, PNC Bank and Wells Fargo.

4 Counter Techniques for DDoS in Cloud

Software Defined Networking (SDN) has showed itself to be a spine in today's network scheme and became a good industry standard. SDN allows us to program and trace the networks and it also assistances the mitigation of some major network glitches. Distributed denial of service (DDoS) attack is the major concern for this. The SDN is concurrently doing the subsequent two tasks. First in a network it Block the malicious flow and second it notify the adjoining networks for a current attack. This way we can avoid the DDoS attack. The other approach intrusion detection algorithms used for DDoS attack detection. This is the model which proposed a system that correlate or mix signature-based IDS using anomaly detection system, which can further leads to achieve high accuracy of the system with IDS. Another approach to mitigate the DDoS

attack is to use of multilayer fair queue that work on priority mapping on a network with traffic deviation which further can directs normal packets will be processed with high priority as well as intruder's packets will be processed with low priority thereby mitigating combined DDoS attack. Furthermore If you are a cloud service provider and need to confirm customers that they can transfer their workloads which are virtual without demanding planning, then SDN is the solution for it. SDN helps in reducing the complexity of the current networks as well as helps to host millions of virtual end to end network without using the methods like VLAN.SDN also enables network administrators to manage network services from a central management tool by virtualizing physical network connectivity into logical network connectivity. So by using the above DDoS attack Detection and prevention techniques we can also implement software defined network which are based on cloud an can also apply some DDoS Prevention based technique like DaMask and furthermore we can also use the statistical based defense system like SDMN which over all work as flow guard for a network (Fig. 2).

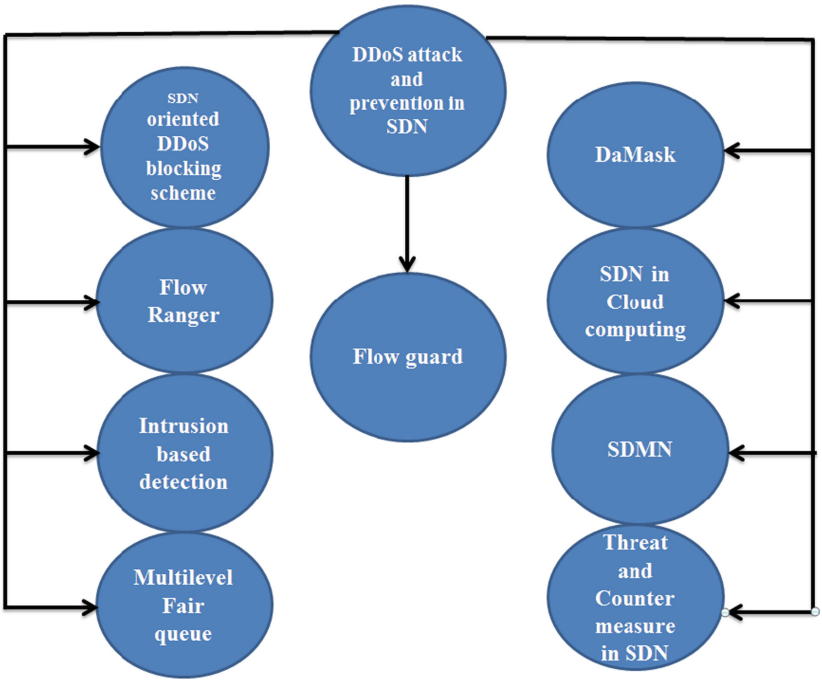

Fig. 2. Counter techniques for DDoS

Numerous studies have been done in the area of DDoS attack defense. One of these studies has investigated the capability of firewalls to mitigate DDoS attacks in the cloud [2]. This empirical study concluded that both software based and hardware-based

firewalls are not enough to defend against DDoS. Thus, more DDoS mitigation strategies are required. Some of them have been presented in a previous paper [7].

This strategy depends on allocating resources dynamically. When DDoS attacks are detected, customers are given additional intrusion prevention servers (IPS) to mitigate the attack. These extra resources are returned to the available resource pool when the attack ends. One more method that has been used to detect attacks is entropy-based. Entropies of some selected meaningful traffic features are measured to detect DDoS attacks. Furthermore, Snort, which is a signature-based detection method, is effective to detect known attacks, but it is less so when it comes to a new attack because the signature was unknown when the attack happened. In addition to the aforementioned defense strategies, anomaly-based methods are considered strong approaches to detect DDoS attacks. The performance of several supervised and unsupervised machine learning algorithms in detecting DDoS attacks are evaluated in [10].

Furthermore, the uses of semi-supervised algorithms to enhance the classifier's intrusion detection performance are used. Authors in [12] have proposed a machine learning-based DDoS attack defense mechanism that is based on analyzing the gathered information from servers' hypervisors and virtual machines. Their method is applied close to the attacker location in the cloud environment. In fact, Neural Networks algorithms are used in several DDoS detection mechanisms. In [12], a hybrid neural network technique that archives high accuracy in detecting DDoS attacks was proposed. A Multi-Layer Perceptron Neural Network was also selected as a base for attack detection methods. NIDS, which is an attack classification method and uses a 2-layered feed-forward neural network is used. In addition to the mentioned algorithms, a Radialbasis- function Neural Network is the core of other DDoS detection mechanisms [12, 13]. Like Neural networks, Naive Bayes algorithms are also used to present accurate defense techniques against network attacks. Furthermore, decision trees are used in many methods to detect attacks. ENDER is a mechanism that applies a decision tree algorithm to detect HX-DoS attacks that combine HTTP and XML messages to target cloud services. Besides utilizing one supervised machine learning classifier to provide network attack defense mechanisms, multi classifiers are combined in one attack recognition method to enhance detection accuracy.

Various studies have evaluated different machine learning classifiers based on their performance in detecting DDoS attacks. Some of them have compared classifiers that belong to many machine learning algorithm types, while other research focused on classifiers located under one machine learning algorithm type. The NSL-KDD dataset was used to compare C4.5, Naive Bayes, Multilayer Perceptron, SVM and PART classifier models in [10] and Bayes Net, Logistic, IBk, JRip, PART, J48, Random Forest, Random Tree and REP Tree in [11]. Additionally RBP, SVM, K-Nearest Neighbor, Decision Tree, and KMeans techniques in [12]. In addition to CAIDA, the DARPA scenario specific dataset and CAIDA Conficker datasets were used to evaluate Naive Bayes, Multi-Layer Perceptron, IBK, R BF network, Bayesnet, J48, Bagging + Random Forest, Voting, Random Forest, and Adaboost + Random Forest to detect DDoS attacks. Moreover, a comprehensive study of existing DDoS attack defense mechanisms has been done, and the authors advocate for the creation of comprehensive, collaborative, and distributed defense mechanisms.

In a network of computer there is the need for network isolation in a private cloud, for that we further identified connectivity of Internet users with private network users. There might be many other sub-categories for the network users, which can be based on functional areas (like R&D) and on the basis of the nature of the service itself or on information sensitivity. The division now is that with SaaS, end users may not directly use the infrastructure while they are using Paas/SaaS. So the cloud needs to accomplish connectivity and strong isolation should be done.

Further the system accepts and processes the multilevel inputs and the aware system finally detect that the attack happened or not. Then system will separate the victim of DDoS attack and concurrently processes the traffic based anomaly detection and drop them (Fig. 3).

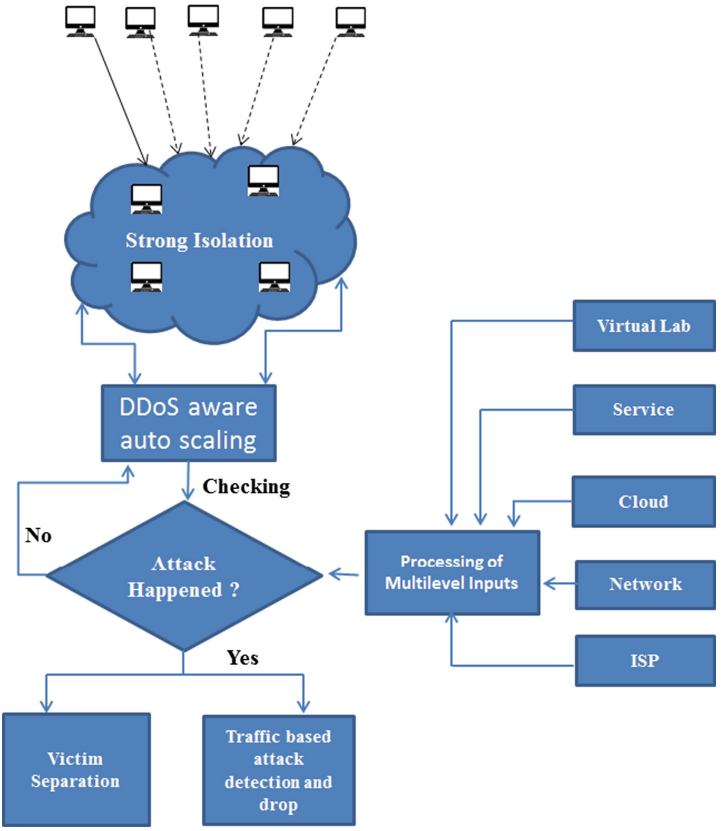

Fig. 3. DDoS aware system

5 Machine Learning for DDoS in Cloud Computing

This research work mainly highlighting on machine-learning methods for detection of Distributed Denial of Service attacks in cloud. To discover and reduce the Distributed Denial of Service attacks in the cloud, many strategies from different security approaches have been presented. One promising detection approach is machine-learning-based. Getting the help of a machine's intelligence enhances analysis and detection accuracy [7].

6 Proposed Model

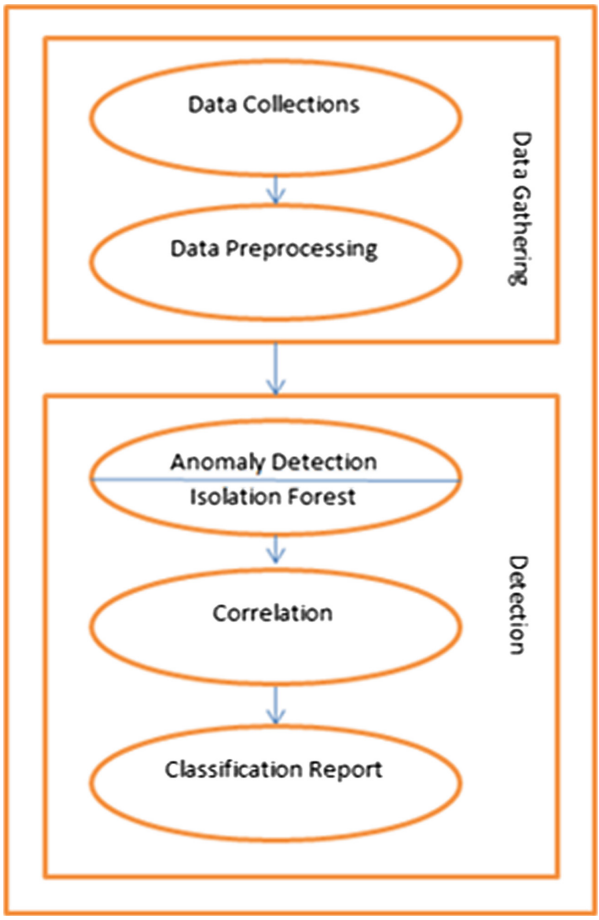

Fig. 4. Proposed model for detection of distributed denial of service attack by use of machine learning technique

As shown in Fig. 4, Data gathering module processes will capture the data packets in a particular format for removing redundant information that has very lower correlation with the detection.

Further the anomaly detection, is referred to the recognition of events or items that do not match to a pattern or to different items present in a dataset, and then it will apply the machine learning procedure on this dataset, By this we conforms that the user is legitimate or not.

We are having the following objectives:

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

6.1 Data Gathering

Here we follow two rules first we collect the data and then we applied Data preprocessing:

6.1.1 Data Collection

This step involves the collection of data or preparing the data.

6.1.2 Data Preprocessing

This is a method that transforms raw data into a logical format. The data of Real-world is habitually inconsistent, inadequate or missing in certain behaviors or styles, and is possible to comprise many errors. This is a said to be great method for resolving such issues.

6.2 Detection

The Anomaly-based Detection (AD) is in recent trends, Because of its ability of identifying novel attack, it has concerned many researchers. We can define the network behavior by detection technique. After that it will generate the result in the anomaly detection. By the specifications of the network administrators the recognized network behavior can be learned or prepared.

AD is centered on a host or a network. Many different techniques are used centered on the type of processing which is related to the behavioral model. Following are the Some of the techniques like Operational or threshold metric model, Statistical based, Statistical Moments or mean and standard deviation model, Time series Model, Genetic Algorithm model, Finite State Machine, Computer Immunology based, Multivariate Model, Cognition based, Adept System Model, Model, Description script Model, Machine Learning based, Baysian Model, Fuzzy Logic Model, Neural Network Model, Outlier Detection Model, User Intention based [14].

6.2.1 Isolation Forests Technique

It is the most recent techniques to detect anomalies. It is based on data point's anomalies which are few and different and are susceptible to a mechanism called isolation.

It uses isolation as an efficient and effective means to detect anomalies. This method requires small memory and has low linear time complexity. It develops model with low number of trees by dividing samples into fixed size data set.

Isolation forest technique isolates the measurements or observations by arbitrarily choosing a feature and then arbitrarily choosing a partitioned value between the minimum and maximum values of the certain feature. It is uncomplicated because it trails few conditions to separate those cases from the usual interpretations. Consequently using the amount of conditions required to isolate a given observation, an anomaly score can be calculated.

The method by which the algorithm constructs the separation is first it creating isolation trees, or random decision trees. After that, for isolate the observation the score is calculated as the length of path.

6.2.2 Correlation Between Data

Generally speaking, when we talk of 'correlation' between two variables, we are referring to their 'relatedness' in some sense. Correlated variables are those which contain information about each other. The stronger the correlation, the more one variable tells us about the other.

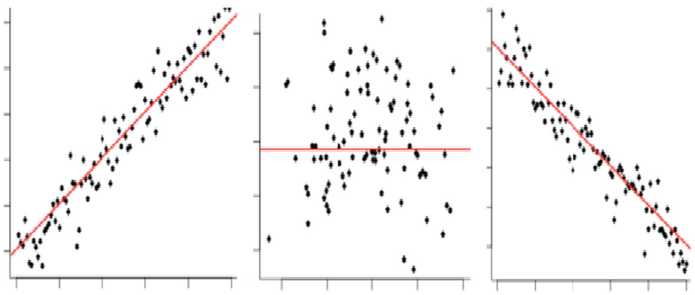

Fig. 5. Correlation

Pearson's Correlation Coefficient (PCC) is a broadly used linear correlation. In mathematically terminology, it is defined as "the covariance between two vectors, normalized by the product of their standard deviations".

The covariance between two paired vectors is a measure of their tendency to vary above or below their means together. That is, a measure of whether each pair tends to be on similar or opposite sides of their respective means.

$$Cov(x, y) = \sum_{i}^{N} \frac{(x_i - \bar{x})(y_i - \bar{y})}{N - 1}$$

The covariance is calculated by taking each pair of variables, and subtracting their respective means from them. Then, multiply these two values together.

If they are both above their mean (or both below), then this will produce a positive number, because a positive × positive = positive, and likewise a negative × negative = positive.

If they are on different sides of their means, then this produces a negative number (because positive × negative = negative).

Once we have all this value calculated for each pair, sum them up, and divide by n − 1, where n is the sample size. This is the sample covariance.

If the pairs have a tendency to be on the same side of their respective means, the covariance will be a positive number. If they have a tendency to be on conflicting sides of their means, the desired covariance will be a negative number. The stronger this tendency, the larger the absolute value of the covariance.

If there is no overall pattern, then the covariance will be close to zero. This is because the positive and negative values will cancel each other out.

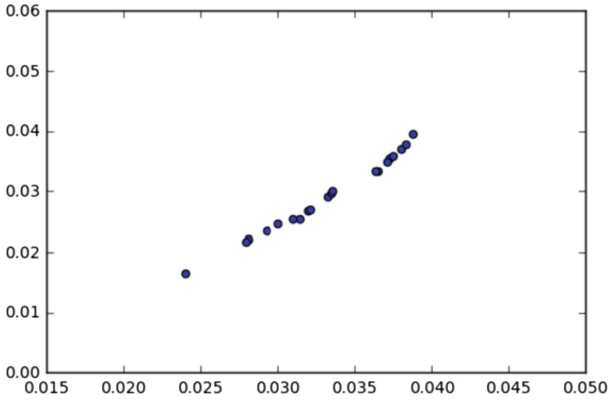

Fig. 6. Covariance

At first, it might appear that the covariance is a sufficient measure of 'relatedness' between two variables. However, take a look at the graph below:

To obtain a more meaningful figure, it is important to normalize the covariance. This is done by dividing it by the product of the standard deviations of each of the vectors.

$$\rho_{xy} = \frac{Cov(x, y)}{\sigma_x \sigma_y}$$

To obtain a more meaningful figure, the normalization of the covariance is very important. This is done by we can divide the covariance by the multiplying of the each vector's standard deviations. The reason this is done is because the standard deviation of a vector is the square root of its variance. This means if two vectors are identical, then multiplying their standard deviations will equal their variance.

Funnily enough, the covariance of two identical vectors is also equal to their variance.

$$Cov(x, x) = Var(x)$$

Therefore, the maximum value the covariance between two vectors can take is equal to the product of their standard deviations, which occurs when the vectors are perfectly correlated. It is this which bounds the correlation coefficient between −1 and +1.

7 Implementation and Result

This proposed model has been implemented on Python 3.6 on Anaconda 5.0.1 with Spyder IDE. We have used DARPA Dataset. The dataset is having malware type of Distributed Denial of Service attack traffic and background traffic that exploits a number of cooperated local hosts (within 172.20.0.0/16 network). The above mentioned compromised local hosts were used to yield a malware DDoS attack on a non-local target.

In order to test the system's ability to deal with DDoS attacks, this article through the open source software simulates the large data traffic DDoS attack, and starts the detection system to detect and address it. The Experimental design is to launch the attacks on the Web Service that has set up the DDoS attack detection system and the

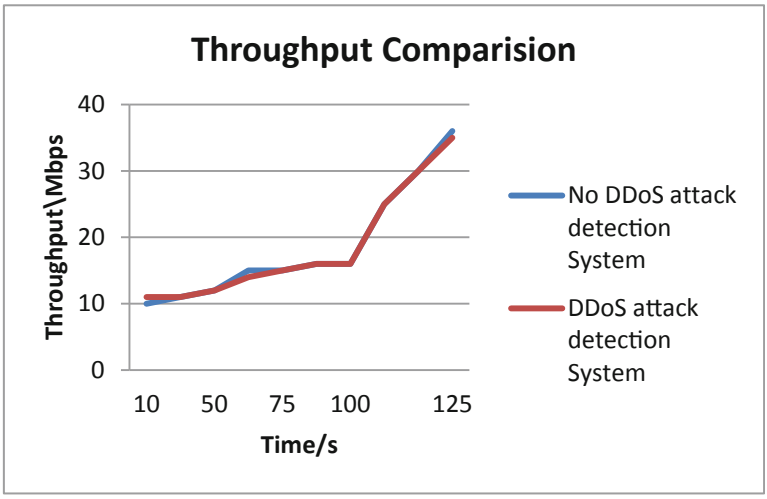

Fig. 7. Throughput comparison

web service that did not build the DDoS attack detection system respectively. The actual impact of DDoS attacks on the server is determined by calculating the Web Service real-time throughput and CPU utilization rate. The final experimental statistics are shown in Figs. 5 and 6. As is shown in Figs. 5 and 6, the throughput of the server increases rapidly after the DDoS attack within 100 s, after which, the throughput of the server in experimental group 1 without DDoS detection system falls sharply with CPU occupancy rate close to 100% whereas that of the server in experimental group 2 with DDoS detection system remains at normal level. Thus, it is proved that Web Service without detection system cannot continue to provide the normal service while the one with the detection system still can operate normally when confronted with DDoS attacks (Figs. 7 and 8).

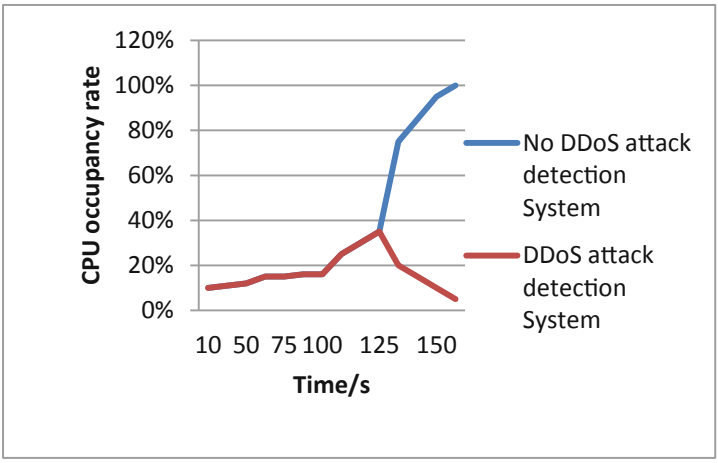

Fig. 8. CPU occupancy rate comparison

8 Conclusions

In this paper we proposed a model which describes the foundations of the main AD technologies and their operational architectures using machine learning along with a classification based method. This method is based on the kind of processing that is associated to the "behavioral" model for the system which is targeted. The noteworthy open issues regarding AD systems are recognized, up on which assessment is given particular prominence. This has been implemented for cloud computing and found the expected results. The presented model establishes an significant idea in the field of IDS to start for addressing Research & Development. In future this proposed model can be used in different computing architectures also.

References

1. Bahrololum, M., Khaleghi, M.: Anomaly intrusion detection system using hierarchical gaussian mixture model. IJCSNS Int. J. Comput. Sci. Netw. Secur. **8**(8), 264–271 (2008)
2. Shawish, A., Salama, M.: Cloud computing: paradigms and technologies. In: Xhafa, F., Bessis, N. (eds.) Inter-Cooperative Collective Intelligence: Techniques and Applications, vol. 495, pp. 39–67. Springer, Heidelberg (2014). https://doi.org/10.1007/978-3-642-35016-0_2
3. El Kafhali, S., Salah, K.: Stochastic modelling and analysis of cloudcomputing data center. In: 20th Conference on Innovations in Clouds, Internet and Networks (ICIN), pp. 122–126. IEEE (2017)
4. Hu, J., Yu, X.: A simple and efficient hidden markov model scheme for host-based anomaly intrusion detection. IEEE Netw. J. **23**, 42–47 (2009)
5. Solanki, K., Dhankar, A., et al.: A review on machine learning techniques. Int. J. Adv. Res. Comput. Sci. **8**(3), 778–782 (2017)
6. Nakkeeran, R., Albert, T.A., Ezumalai, R.: Agent based efficient anomaly intrusion detection system in ad-hoc networks. IACSIT Int. J. Eng. Technol. **2**(1), 52 (2010)
7. Zekri, M., El Kafhali, S., Hanini, M., Aboutabit, N.: Mitigating economic denial of sustainability attacks to secure cloud computing environments. Trans. Mach. Learn. Artif. Intell. **5**(4), 473–481 (2017)
8. Xiao, P., Qu, W., Qi, H., Li, Z.: Detecting DDoS attacks against data center with correlation analysis. Comput. Commun. **67**, 66–74 (2015)
9. Ahmed, A.A.E., Traore, I.: Anomaly intrusion detection based on biometrics. In: IEEE Workshop on Information Assurance (2005)
10. Sheta, A.F., Alamleh, A.: 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, p. 15 (2015)
11. Bhuse, V., Gupta, A.: Anomaly intrusion detection in wireless sensor networks. ACM J. High Speed Netw. **15**, 33–51 (2006)
12. Shirazi, H.M.: Anomaly intrusion detection system using information theory, K-NN and KMC Algorithms. Aust. J. Basic Appl. Sci. **3**(3), 2581–2597 (2009)
13. Yang, D., Usynin, A., Hines, J.W.: Anomaly-based intrusion detection for SCAD systems. In: IAEA Technical Meeting on Cyber Security of NPP I&C and Information systems, Idaho Fall, ID (2006)
14. Manikopoulos, C., Papavassiliou, S.: Network intrusion and fault detection: a statistical anomaly approach. IEEE Commun. **40**, 76–82 (2002)