MAINTAINING CLOUD PERFORMANCE UNDER DDOS ATTACKS

Moataz H. Khalil[1, 2], Mohamed Azab[2], Ashraf Elsayed[3], Walaa Sheta[1, 2], Mahmoud Gabr[3] and Adel S. Elmaghraby[1, 2]

[1]CECS Department, University of Louisville, Kentucky, USA
[2]The City of Scientific Research and Technology Applications, Egypt
[3]Department of Mathematics & Computer Science, Faculty of Science, Alexandria University, Alexandria, Egypt

ABSTRACT

The popularity of cloud computing has been growing where the cloud became an attractive alternative rather than classic information processing system. The distributed denial of service (DDoS) attack is one of the famous attacks to cloud computing. This paper proposes a Multiple Layer Defense (MLD) scheme to detect and mitigate DDoS attacks which due to resource depletion. The MLD consists of two layers. The first layer has an alarm system send alarms to cloud management when DDoS attacks start. The second layer includes an anomaly detection system detects VM is infected by DDoS attacks. Also,MLD tested with a different DDoS attack ratio to show scheme stability. MLD evaluated by The energy consumption and the overall SLA violations. The results show the great effect of the MLD to reduce the energy consumption and the overall SLA violation for all datasets. Also, the MLD shows acceptable stability and reactivity with different DDoS attack ratio.

KEYWORDS

Cloud Computing, Energy consumption, Service Level Agreement, DDoS attack, anomaly detection, Availability.

1. INTRODUCTION

A pay-as-you-go (PAYG) model is an innovative paradigm was designed by cloud computing providers to apply for application, platforms, services and computing resources to users. [1]. various Quality of Service (QoS) aspects, like performance, availability, and reliability are used to measure performance of different services provided by cloud computing platform Theseperformance metrics are explained in a Service Level Agreement (SLA) negotiated between users and cloud providers. Cloud services are classified as service as Software as a Service (SaaS), Platform as a Service (PaaS), and Infrastructure as a Service (IaaS). SaaS is a service of software deployment where a service is hosted as a service delivered to users across the Internet. SaaS is used to mention to business software rather than user software, which belongs to Web 2.0. without needing to install and execute a service on a user‘s computer it is considered as a way for businesses to get the same profits as commercial software with smaller cost outlay.

DOI: 10.5121/ijcnc.2019.11601

The cloud computing provider suffers from a lot of stripes results from growing pressure from deliver services where platform as a services is a form of cloud computing that enables potential to assistant developer's designs, write, and test web applications services that presented to customers.a lot of venders such as Salesforce.com (Force.com), Microsoft (Azure, starting next month), and startups such as Wave Maker are contributed in arising up online development environment.One development language or methodology has been used by these platforms which is good thing for the enterprise.

A high level of availability and reliability of the application and services have been available where the cloud computing environment is characterized by a high volatility. the ability of system to perform as possible when the services are requested where it is not failed or repaired action it is called availability. Cloud computing reliability is defined as the framework of security or the framework of resource and service failures. the ability of components and parts of system to do the required jobs for section of time with a certain level of confidence it is defined as reliability.

The relation between the reliability and availability controls with the third compound called maintainability. Performing a successful repair action within a certain time it is named maintainability.Reliability, Availability, and Maintainability (RAM) encompass the essential features of SLA. RAM are correlated in such as a way that it is necessary to have both high reliability and good maintainability in order to achieve high availability [2]. The relation between reliability and availability isa positive relationship at constant maintainability.

One of the most important features of the cloud is high availability service to the customer. Cloud computing focuses on that user can get information anywhere anytime. Availability does not only refer the software and data but also it provides hardware as demand from authorized users. The availability of cloud computing services is targeted by cloud resource based attack such as Denial of Service (DoS) or Distributed Denial of Service (DDoS) attack [3,4]. The DDoS attack is looked one of the greatest serious attacks in the cloud environment.DDoS is known as a cyber-attack where the attacker tries to make a machine resource unavailable to its intended users by momentarily or forever disrupting services of a host connected to the internet. DDoS defines as a group of machines that are targeted at a particular service. DDoS attacks goal to consume a system's resources such that it compromises its capability to offer the intended service and thus rendering it unreachable [5].

Many DDoS attacks cases gained a lot of attention in the research community. Such as, lizard Squad attacked cloud-based gaming services of Microsoft and Sony which removed down the services on Christmas day in 2015. A massive DDoS attack was targeted The cloud service provider Rack space on its services. Amazon EC2 cloud servers faced a massive DDoS attack. The DDoS attack caused heavy downtime, business losses and many long-term and short-term effects on the business processes of victims [6]. The economic losses per hour at peak times is 470% more than the former year. In March 2015, Greatfire.org was targeted by a heavy DDoS attack by costing it an enormous bill arrive at 30000$ daily on Amazon EC2 cloud [7]. In [8], the authors reported that up to 444000 USD as totally is the average financial damage by a DDoS attack. The categories of DDoS is classified as bandwidth based and resource-based attacks. The bandwidth attack devours the bandwidth of the victim or target system by overflowing with the undesirable traffic to stop legitimate traffic from arriving the victim network. The bandwidth is divided into flooding attack and amplification attack. On another side, the resource depletion targets to exhaust the victim system's resources. The main two branches of the resource depletion attack are protocol exploit attack and malformed packet attack [9].

The DDoS attacker harnesses the most important advantages of cloud computing like pay-as-you-go model, auto-scaling, and multi-tenancy to get its goal. For pay-as-you-go model, the cloud instances are rented on an hourly basis and thus the minimum renting period is an hour. A virtual machine (VM) owner may need to update its own resources on-the-fly as and when required. In addition, cloud computing offers better hardware utilization, a consumer does not want provisions like power, space, cooling, and maintenance. Pricing or accounting plays a vital role in DDoS attacks in the cloud. Attackers need only to pay the cost of hours that VMs are active [6]. Multiple providers support the auto scaling concept [6, 10]. For auto-scaling, this property permits allocation of additional CPUs, memory, storage, and network bandwidth to a VM when the resources are required or removed from a VM when a VM does not need these resources. In addition, it can also transfer from the host to another host. The advantage of multi-tenancy gives the benefit that is an architecture in which a single instance of a software application serves multiple customers. A DDoS attacker uses this advantage specifically such that if attacker success to attack single VM which serves a lot of applications, a lot of applications will be out of order [6].

The DDoS attacks of effects are categorized into two sets; direct and indirect effect [6, 11, 12, 13]. The direct attack effect such as service downtime, auto-scaling driven resource/economic losses, business and revenue losses, economic losses due to the downtime, and the service's downtime which is based on the victim service. while indirect attack effect on the cloud such as energy consumption costs, reputation and brand image losses, attack mitigation costs, collateral damage to the cloud components and the effects due to recent smoke-screening attacks. The economic losses have multiple phases in direct and indirect DDoS effect. as Economic Denial of Sustainability (EDoS) attack is the economic losses depend on DDoS attack which is known or Fraudulent Resource Consumption (FRC) attack. The DDoS attack takes the shape of an EDoS attack when the victim service is hosted in the cloud [6, 11].

The defense System for DDoS attacks in the cloud is categorized into two main modes: proactive defense mode and reactive defense mode. The proactive defense mode includes DDoS attack prevention. The reactive defense mode includes attack detection and attack mitigation and recovery [6, 9]. At the proactive defense mode, the DDoS attack prevention methods are based on one or more functions like; challenge response, hidden servers or ports, restrictive access, and resource limit.

For the reactive defense mode, The DDoS attack detection method works in a situation that the attack has been done. The DDoS attack detection method starts to run according to signals that are sent from the cloud management system. The received signals announce that the attack starts. Also, cloud performance will degrade. The DDoS attack detection is constructed based on one or more functions like; anomaly detection, sources and spoof trace, count-based filtering, botcloud detection, resource usage. Anomaly detection defines as the process of recognizing unanticipated patterns or events in datasets, which are varied from normal patterns or events. Anomaly detection is classified into three groups. They are; supervised, semi-supervised, and unsupervised anomaly detection. Supervised anomaly detection is similar to supervised methods. So, the labeled train and test data are required. On another hand, the unsupervised anomaly detection is suffered from highly sensitive to outliners.

For the attack mitigation and recovery, the proposed methods support an infected server to remain serving requests in the presence of an attack. The published methods are based on one or more functions like; victim migration, OS resources management, Software Defined Networking (SDN), DDoS mitigation as a Service. In Cloud computing, it is very vulnerable to a DDoS attack due to the structural features of the Cloud system. When logical resources have been being delivered virtualization layer on physical resources, a set of virtual machines can be affect by DDoS attacks on one physical resource which virtual machines used the resource on physical resources between them. DDoS attack consumes the system resources, and users can not able to receive reliable services. As a result of DDoS attacks due to SLA violation. At the same time, DDoS attack depletes the system resources which causes a high consuming for CPU/memory resources. CPU/ memory resources are two of the highest compounds causing energy consumption. So, the DDoS attacks have an indirect effect on energy consumption by raising CPU/memory utilization resources that become them busy all the time.

This paper proposes Multiple Layer Defense (MLD) scheme. The proposed scheme sends alerts to a cloud management system for notifying about the attack. Also, the proposed scheme mitigates the effects of DDoS attack on the cloud system. The proposed defense scheme focuses on resource depletion DDoS attack. The proposed scheme consists of two layers. The first layer aims to alert the cloud system when attacks start. This layer is based on a prediction method. The prediction method depends on a rigid regression learning model. The prediction method works to forecast the requested workload size by the users for cloud services. The predicted requested workload size is based on the requested workload size in the previous stage, day, and time. The predicted requested workload size uses as a dynamic threshold for the workload size. If the amount of the real requested workload size is larger than the predicted requested workload size, the cloud manager announces that the cloud exposes to attack.

To avoid the false positives and false negatives alarms can produce from the first layer, the second layer aims to detect anomaly patterns which are presented on the DDoS attacks. The detection process depends on the behavior of the VM resource utilization during the lifetime of VM. The second layer contains a one class support vector machine model to detect the DDoS attacks. A most of published papers studied the effect of DDoS attacks on cloud computing through response time. This paper will gauge the effect of DDoS attack in cloud computing through the SLA violation and energy consumption. The key contributions of this paper are:

1- For predicting the size workload in cloud computing, the prediction model based on rigid regression is proposed.
2- A new dynamic threshold for the number of jobs requested proposes is proposed.
3- Using SLA violation and energy consumption to evaluate the performance of the cloud with the MLD scheme under DDoS attacks.

The organization of paper is as follows: section 2 has a set of research manuscript published. Section 3 consists of an explanation and analysis of the dynamic threshold for the workload size. Section 4 discusses and analyzes the performance SVM-one class. Section 5 explains the process of clustering workload. Section 6 has the experiment design and metrics evaluation. in section 7, The results and analysis are explained. Conclusion and future work are discussed in section 8.

2. RELATED WORK

According to previous sections, the defense System classes for DDoS attacks in the cloud are; attack prevention, attack detection, and the attack mitigation and recovery. This paper is going to focus on the DDoS attack detection and mitigation with more focus on approaches based on count based filtering, resource usage methods, and OS level resource management methods. The next section is going to discuss a set of the published researchers in cloud computing workload prediction area.

2.1. Workload Prediction in Cloud Computing

In this section, the proposed workload size prediction approaches will discuss. To develop a more accurate prediction approach, data mining and machine learning techniques including regression, decision trees, neural networks (NNs), fuzzy logic, genetic algorithm (GA), support vector machine (SVM) that apply to predict workload submitted in cloud computing. A new service cloud architecture is presented and a linear regression model was applied to predict the workload size trace in [14]. The workload prediction aimed to forecast the size of services requested at the next time interval. The authors used a linear regression model (LRM) to solve this problem and compared with other models like autoregressive moving average method filter (ARMA), mean workload prediction, and max workload prediction. The LRM outperforms other methods. The Workloads are used in this paper consists of video service for 6 hours every time interval in a service cloud. A set of published papers on workload prediction, the different methods of regression like linear regression and Auto Regression (LR) Integrated Movie Average (ARIMA) are mostly used to predict the size of request workload in the cloud computing. The main shortage of LR that it does not suitable for the cloud computing environment. The cloud computing environment has a high change in the size of the request workload. In addition, another the disadvantage of the linear regression, the ARIMA model selection process based on greatly on the competence and knowledge of the scientists to yield targeted results [15].

A Cloud Resource Prediction and Provisioning scheme (RPPS) is proposed in [16]. The RPPS proposed for automatically estimate the future request and achieve proactive resource provisioning for cloud applications. RPPS is based on an autoregressive integrated moving average model (ARIMA) to forecast the workloads in the near future, combines both coarse-grained and fine-grained resource scaling under various situations, and adopts a VM-complementary migration approach. RPPS can resolve a predictive resource provisioning challenge when enterprises confront request variations in the cloud data center. The model of RPPS assessed with traces collected by the authors using typical CPU intensive applications and as well as workloads from a real data center. Anew prediction approach is suggested in [17]. The suggested prediction approach classifies the workload and assigns diverse prediction models according to the workload features. The key idea is that the authors convert workload classification into a 0–1 programming problem, and formulate an optimization problem to maximize the prediction precision, and then present an optimization algorithm. The proposed approach tested with real traces of typical online services to evaluate prediction method accuracy.

In [18], the authors extended the research of [17]. The authors discussed the classified prediction approach in the perspective of the IaaS layer. The authors analyzed the problems in a large-scale heterogeneity cloud environment and assess the classified prediction method with google cluster real trace data. An adaptive categorical prediction scheme is proposed according to factors in the IaaS layer. The suggested approach categorizes the workloads into different sets corresponding to

different prediction methods as in [17]. In addition, in order to modify the prediction scheme timely, feedback from the workload monitoring was applied. With establishing an integer programming model, an ideal method has been adopted to classify workloads.

Three models to predict the workload based on analyzing monitoring data are proposed in [19]. The first prediction model uses a time series approach to analyze monitoring data. The authors compared a set of time series approaches. The time series approaches include Movie Average (MA), Auto Regression (AR), ARIMA, Difference Model (DM) and median model (MM). The time series approach can analyze and predict the CPU utilization using the history data, but sometimes it lost the real data and has low prediction accuracy. A Kalan filter model is proposed as the second prediction model to forecast the cloud workload. Kalman filter model can estimate the true data based on the observed data. The Kalman filter model works in a two-step process, including the prediction step and update step. The pattern matching model is the third prediction model. The pattern matching works by matching the sequence with some history patterns. The third model is based on the string matching algorithm and Euclidean distance. This model includes two steps, preprocessing and match. All prediction models have been evaluated by the Mean of Absolute Percentage Error (MAPE). The results showed that the Kalman filter has the least MAPE compared with all prediction models.

In [20], the realization of a cloud workload prediction module for SaaS providers is presented. The proposed prediction model is based on the autoregressive integrated moving average (ARIMA) model. The accuracy of future workload prediction evaluated using real traces of requests to web servers. In addition, the effect of the achieved accuracy in terms of efficiency in resource utilization and quality of services QoS assessed. A Bayesian model is proposed in [21] to predict virtual resource requirement of applications in short and long-term. The factors considered for prediction in the model are a day, weekday or weekend, time-interval of application access, workload, benchmarks, and availability of virtual machines, etc. Dependencies between related parameters were identified. The assessment of the model is carried out using the cross-validation method on the basis of training, validation and test datasets. The datasets are reflected CPU intensive transactional requests. The SamIam Bayesian network simulator is used to build the model. The presented model verified cross workload traces of Amazon EC2 and Google CE data centers in real time scenarios. The main limitation of previous approaches focuses on static, and data size. A lot of papers ignores the dynamism environment of cloud computing. The data size used in the previous approaches is less than the real data size of cloud computing. In addition, the accuracy of the proposed approaches still low.

2.2. DDoS Attack Detection & Mitigation

In this section, the published papers proposed detection and mitigation of DDoS attacks methods will discuss. In [22], the authors proposed a novel detection and mitigation technique against EDoS attack in cloud computing called EDoS-Shield. The EDoS-Shield aims to confirm whether the user requests are a legitimate person or generated by bots. The EDoS-Shield has two lists, the white list for a legitimate person while the blacklist for the un-legitimate person (DDoS attack). The detection mechanism works by forwarding the first request to a verifier node in EDoS-Shield architecture. This verifier node is responsible for the detection process and updating the white and blacklists based on the results of this detection process. The following requests sending from the bots will be obstructed by a virtual firewall where their IP addresses will be assigned in the blacklist. On the other hand, the following requests scheduling from legitimate users will be send directly to the target cloud service where their IP addresses will be placed in the whitelist. In [23],

the authors extended the previous work published in [22]. The issue of IP spoofing discussed within EDoS-Shield architecture to detect and mitigate the DDoS attacks. In the enhanced EDoS-Shield architecture, the authors applied the time-to-live (TTL) value found in the IP header for the objective of detecting the IP spoofed packets. To mitigate, these spoofed packets will be deleted before reaching the protected server. In this architecture, when a V-Node achieves detection of a request, it will assign to the corresponding TTL value related to the source IP address. With enhanced EDoS-Shield architecture, both values of the IP address and TTL value will be allocated in the white or blacklists. In the future time, the acquiring information will be help to distinguish the packets having spoofed IP addresses, and then can selectively filter out these packets through a virtual firewall (VF).

In [24], the authors used the request count threshold idea for detecting DDoS attacks. The threshold is based on the basis of human behavior at the time period. For mitigation, all subsequent requested are dropped all from the same IP for a finite period. The detection and mitigation system is proposed in [25]. The defense system includes a virtual machine monitor and an isolated system. The job of a virtual machine monitor (VMM) detects the DDoS attacks depends on the resources consumed. Once the DDoS attack detects, the VMM creates an isolated environment for running application by using duplication. When the isolated environment is completed, the isolated environment does not communicate with others anymore, because it has no I/O function. The isolated environment simply keeps the execution of tagged applications. After the DoS attack stops, VMM puts back the OS status as well as the tagged applications in the isolated environment to the VM and the isolated environment can be shut by VMM.

A dynamic resource allocation strategy is proposed in [26] to prevent DDoS attacks against individual cloud customers. The DDoS attack occurs when the time of a packet spends in non-attack mode (constant) is larger than the time of a packet spends in attack mode. The Intrusion Prevention System (IPS) is responded to monitoring the time of packets. To relieve the effect of DDoS attacks, the proposed method will be automatically and dynamically located additional resources from the available cloud resource pool, and a fresh VMs will be replicated depending on the image file of the original IPS using the current replica technology. All IPSs will work together to elect attack packets out and guarantee QoS for benign users at the same time. When the volume of DDoS attack packets decreased, the proposed method will automatically reduce the amount of its IPSs and release the additional resources back to the available cloud resource pool. The amount of IPSs that require to sustain the goal depends on the volume of the attack packets.

The defensive framework is proposed in [27] called ATOM. ATOM applies cross the IaaS layer and provides automated tracking, orchestration, and monitoring of resource utilization for a large amount of VMs running on an IaaS cloud, in an online mode. ATOM presents an online tracking module running at Node Controller (NC) and continuously tracks several performance metrics and resource utilization values for every VMs. The Cloud Controller (CLC) is referred to as the tracker, and the NCs are denoted as the observers. The two main objectives of tracking and monitoring are; (1) exchanging the basic view at the CLC with a realization of system status, with minimum overhead, (2) studying the performance of resource utilization data reported by the online tracking module to discover an anomaly. ATOM includes a naïve method to define a threshold value for any an interesting metric choose by a user. Enhanced ATOM is able to defend the dynamic and complex attacks and anomalies in cloud computing. The optimized ATOM applies a dynamic online monitoring mode developed depended on Principal Component Analysis (PCA) to do mining in the resource data and creates anomaly information to assistance further analysis by the orchestration component when this happens. The orchestration component

in ATOM leverages virtual machine introspection (VMI) tools. A VMI is a process that permits indirect inspection and manipulation of the state of virtual machines. The monitoring component sends the VMI tool with a priori knowledge of what might have gone wrong. Also, it works as a trigger to voice VMI tools when and where to do introspection. With this information, the overhead of using VMI techniques is greatly reduced.

The authors in [28] proposed a DDoS aware resource allocation strategy in which the overloaded VMs are not directly flagged for resource increase. Instead, authors propose to separate the traffic and increase the resources only on the basis of the demands of genuine flagged requests. A set of published papers are used machine learning techniques to detect DDoS attacks. An effective DDoS attacks detection approach based on K-nearest Neighbor traffic classification with correlation analysis (CKNN) is proposed in [29]. The approach benefits from correlation information analysis of training data. With the correlation analysis, the hidden relations can find in the training data from a data center, which able to improve the classification accuracy and is not affected by the density of training data. To reduce the overhead of KNN, the authors map the training data into the grid. The testing examples are only calculated with the training examples in neighboring cells rather than all the training data by applying the r-polling method which can decrease the overhead of CKNN professionally. Furthermore, the CKNN method is affected less by the mass of training data which directly effects the effectiveness and precision of traditional KNN classifier. The proposed approach tested with the Internet, data center traffic trace and the KDD'99dataset.

A DDoS attack detection system is proposed in [30]. The DoS attack detection is based on using Multivariate Correlation Analysis (MCA). MCA has the ability to extract the geometrical correlations between network traffic features which help to gain more accurate network traffic characterization. The presented MCA-based DoS attack detection method used the standard of anomaly-based detection in attack recognition. Which due to the presented solution ableof discovering known and unknown DoS attacks efficiently by knowledge the patterns of legitimate network traffic. A triangle-area-based method is presented to improve and to speed up the process of MCA. The effectiveness of the proposed detection method assessed using KDD Cup 99 data set, and the effects of both non-normalized data and normalized data on the performance of the proposed detection system are observed. A profile based network intrusion detection and prevention system are presented in [31]. The proposed system aims to secure the cloud against malicious insiders and outsiders. The proposed system mix both fine-grained data analysis and Bayesian technique approach to detect DDoS attacks using unsupervised learning algorithm. The goal of the proposed system is detecting network attacks, such as TCPSYN flooding.

The authors in [32] designed three stages of anomaly detection to detect DDoS attacks. The first stage is the monitoring stage, which uses a rule-based system to preprocess known DDoS attack patterns. The second stage offers lightweight anomaly detection. The second stage forecasts the future load on each customer interface using time series modeling. The traffic volume over the network is divided into large and small volumes during the time axis, and the Bayesian technique applied to analyze DDoS attack candidate on the network topology. The last stage is focused on anomaly detection to identify both known and unknown DDoS attack patterns using an unsupervised learning algorithm. The main limitations of the previous approaches are; static, user communication based, and network focused. The static method is not suitable for dynamic cloud computing behavior. If the malicious user success to simulate the behavior ofa real user, the cloud system will be a victim of a huge number of attacks. A lot of the published papers are focused on

detecting DDoS attacks based on network devices behavior. In addition, there are continuously developing on IP spoof trace method which requires more development for detecting DDoS attack cross the network. According to these shortages, the proposed scheme will depend on detecting DDoS attack on the cloud side. The cloud side means that the DDoS attacks are going to detect based on monitoring cloud computing performance. This will be a good help to avoid the challenges of IP spoof. In addition, the proposed scheme does not depend only on the network, the CPU and memory resource utilization will be considered. an approach for visualizing network attacks data using clustering is proposed in [33]. The proposed approach based on K-means algorithm with the Kdd Cup 1999 network data set to evaluate the performance of an unsupervised learning method for anomaly detection. The proposed approach consists of three stages. After entering corrected KDD dataset, the first stage fragments the 37 attacks which are founded in this dataset into four general categories (DOS, Probe, R2L, and U2R). The second stage uses Cluster 3.0 tool for apply k-means technique to cluster attacks. The third stage implements Tree View visualization tool to visualize k-means result. The results of the evaluation showed that a high detection rate can be achieve while maintaining a low false alarm rate.

3. MULTIPLE LAYER DEFENSE (MLD) SCHEME

MLD scheme aims to reduce the harmful effect of DDoS attacks in a cloud computing environment. MLD focuses on reducing SLA violation as a direct effect of DDoS attacks. in this study, SLA violation is caused by reducing availability and reliability. Also, MLD aims to reduce energy consumption as an indirect effect of DDoS attacks. The energy consumption raises up when cloud resources become high utilization. In the next sections, two layers of MLD will be explained.

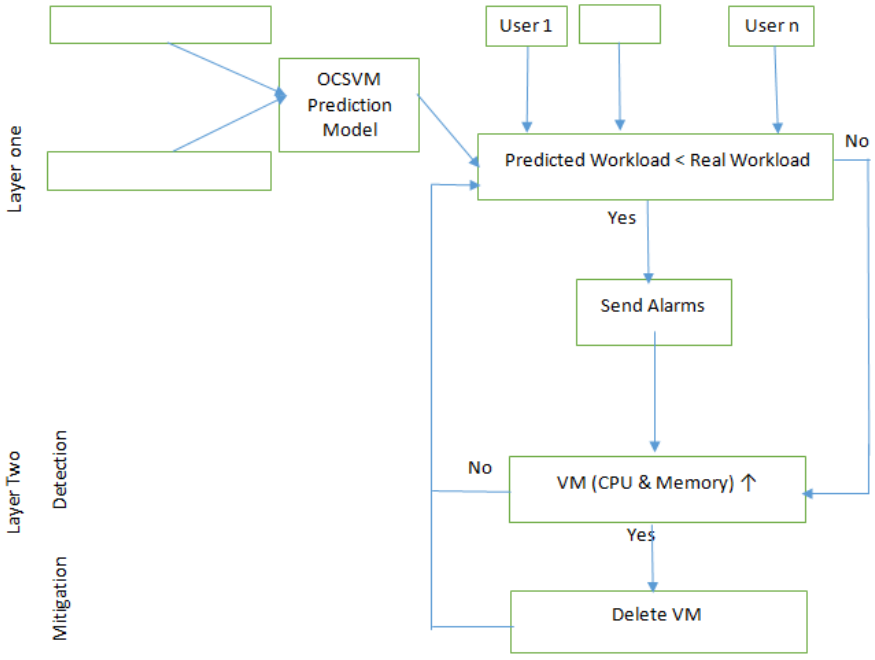


Figure 1. The Proposed MLD Architecture.

Figure 1 explains the MDL architecture. At the first stage, the prediction model receives the input parameters which are data and time of previous workload and the size of the requested workload. The comparison between real workload requested by users and the predicted workload determines that if the alarms will send to cloud management or not. At the second stage, the detection and mitigation will run. In the case of the resources utilization of VM is in an increasing manner. The mitigation process will run by detecting the VM has an increasing manner.

3.1. Layer One: Dynamic Threshold for Workload Size

A lot of published papers are used as a static threshold to detect DDoS attacks. While the environment of cloud computing is very dynamic. This paper proposesa method to establish an alarm system. This alarm system is based on creating a dynamic threshold for the size of request jobs. The comparison between the real jobs requested with dynamic threshold helps to detect DDoS attacks. If the real jobs requested is higher than the threshold, then the alarm system sends alters to cloud management to notify that DDoS attacks start. The proposed method is based on a rigid regression learning model. The rigid regression has two main advantages. First, rigid regression has a penalty term that reduces overfitting. The rigid regression uses L2 penalty term as shown in equation 1. Second, the rigid regression works to increase the correlation between the model features which improves the performance of a model at all. The objective of rigid regression is minimizing the gradient descent function (J). Equation 1 explains the objective of rigid regression.

$$J(\beta) = \sum_{i=1}^{n}(Y_i - \sum_{j=1}^{p}(X_{ij}\beta_j))^2 + \lambda \sum_{j=1}^{p}\beta_j^2 \quad (1)$$

Where Y is observed value,$\sum_{j=1}^{p}(X_{ij}\beta_j)$ is predicted value, β is regression coefficient, λ is shrinking parameter, and$\sum_{j=1}^{p}\beta_j^2$ is L2 penalty term.

The day, hour, and previous request workload size are used as features for the rigid regression learning model. Two processes are applied as a preprocessing operation to increase the learning time and improve model performance. First, the regression performance improves with a low number of features. The hour and day are mixed to be one feature. Second, the normalization process is applied for all features to keep the same range between 0 and 1. The normalization process aims to increase converge speed. The Leave-One-Out cross-validation learning technique was applied to in train stage.

3.2. Layer Two: One Class Support Vector Machine (OCSVM)

The different between the anomaly detection and the classification methods that the anomaly detection is able to use for an unlabeled data, taking only the internal structure of the dataset into account [34]. This paper implementssemi-supervised method where the label data does not require. At the same time, a semi-supervised is less sensitive to outliners. The paper process a method to detect and mitigate DDoS attack. The presented method depends on one class support vector machine (OCSVM) model. The main objective of the proposed method that reduces the cost of the false positive and false negative alarms can send from layer one.

The OCSVM model transforms input data into a high dimensional feature space by applying the kernel and iteratively finds the maximal margin hyper plane which best separates the training data from the origin. The OCSVM might be seen as a normal two-class SVM where all the training

examples lies in the normal class, and the origin is taken as the abnormal class. A separated model created for different resources utilization. The Gaussian kernel is used in OCSVM model. LibSVM library on JAVA used to create a model.

4. WORKLOAD CLUSTERING

K-mean is a well-known unsupervised learning algorithm. The k-Means method allocates N data points to k diverse clusters, such that the clusters number needs to know a priori. This paper uses the K-means for clustering the size of workload requested. The objective from the workload cluster is simulated the real number of VM type requested by users. A different number of clusters are tested starting from 1 to 10. A heuristic approach is applied to determine the best number of clusters. The heuristic approach is called the Sum of Squared Distances (SSD). SSD represents the sum squares of the distance between points and cluster center [35]. The low values of SSD refer to that the cluster is coherent. Figure 2 shows the relation between SSD and the number of clusters. Form figure 2, it clears that the best number of clusters is 4.

5. EXPERIMENTAL DESIGN

The MLD scheme will be evaluated in a practical cloud scenario, the Clouds simulation toolkit has been used. CloudSim is the most popular simulation tool available for the cloud computing environment. It is an event-driven simulator built upon the core of grid simulator GridSim. Base programming language for CloudSim is Java. CloudSim is open source, so its modules are easy to extend based on Java. A set of experiments will test MLD over a different real workload including DDoS attacks. Also, MLD will test under a different ratio of DDoS attacks. The MLD focusses on resources depletion DDoS Attack.

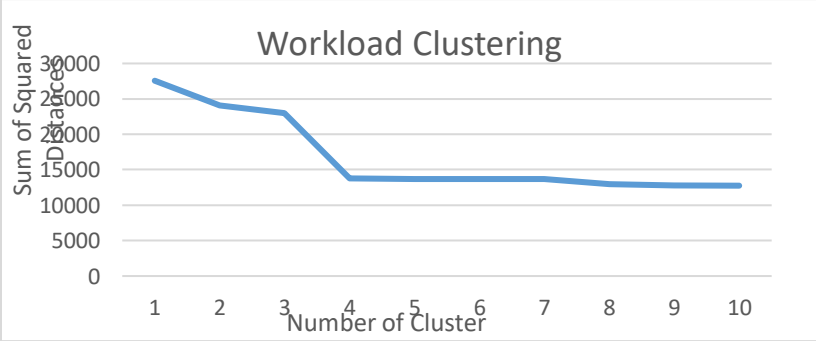


Figure 2. The Relation between SSD and Number of Cluster.

5.1. Experiment Setup

For hardware, the experiments were test cross a data center that has 800 heterogeneous physical nodes. A data centerenvironmentisheterogeneous where has of two types of hosts; the first half of the hosts are HP ProLiant ML110 G4 servers with 1,860 MIPS per core, and the other half are HP ProLiant ML110 G5 servers with 2,660 MIPS per core. Each server has 2 cores, 4 GB of memory and 1 GB/s of network bandwidth. The power consumption of active 2 servers in the simulation is derived from the corresponding figures in the Standard Performance Evaluation Corporation (SPEC) [36]. Table 1 summarizes a data center configuration.

According to the result of the workload clustering, the best number of clusters is 4. So, four different virtual machines are used following Amazon EC2 instances [37]: High-CPU Instance (2500 MIPS, 0.87 GB), Extra Large Instance (2000 MIPS, 1.74 GB), Small Instance (1000 MIPS, 1.74 GB), and Micro Instance (500 MIPS, 613 MB). At beginning the simulation, VMs are hosted according to the resource needsthat is defined by the VMs. Table 2 summarizes virtual machines configurations.

For software, a real-world workload represented VM utilization. The MLD defense is tested with three different days extracted from Google Cluster Data (GCD) real workload. The GCD workload consists of the resources utilization form Google Cluster Data (GCD) dataset for a 29-day period in May 2011 [38]. The GCD workload includes 670983 jobs, each job with one or more tasks with a total number of tasks of 144841618, and contains the normalized value of the average number of used cores and the utilized memory. To create the CPU and the memory utilization of VMs, the tasks of each job was aggregated by summing their CPU and memory consumption every five minutes in a period of 24 hours. We extracted experiment workload form GCD workload by extracting computing jobs (high priority and non-missed value). The computing jobs are high resources utilization in GCD compare with other jobs, such that if a machines resource utilization is very full but 90% of utilization is attributed by low jobs, a machine is considered idle [39]. Table 3 summarizes the characteristics of the workload submitted by three days at different days and hours.

Table 1. Data Center Configuration.

Host	MIPS	Number of Cores	Memory	Bandwidth
HP ProLiant ML110 G4	1860	2	4 GB	1 GB /s
HP ProLiant ML110 G5	2660	2	4 GB	1 GB /s

Table 2. Virtual Machine Configuration.

VM Type	MIPS	Memory
Extra Large	2000	1.74 GB
High-CPU	2500	0.87 GB
Small	1000	1.74 GB
Micro	500	613 MB

5.2. Evolution Metrics

The utilization of resources is measured every 5 minutes over 24 hours which is the system lifetime. The proposed framework evaluates by the following metrics:

1- The Number of host hot during 24 hours.
2- Energy Consumption for all hosts during simulation time.
3- The Number of VM migration.
4- Overall SLA volition.

Table 3. Workload Trace Characteristic.

Name	Day	Hour	Micro Machines size	Small Machines size	High Machines size	Extra Machines size	Total
D1	1	9	1786	603	289	318	2996
D2	18	1	1536	90	18	24	1668
D3	28	17	2310	30	27	6	2373

The power of hosts depends on the maximum power of the hosts and CPU utilization of hosts. The energy consumption is calculated as the difference between host powers for time cascaded. (2) & (3) calculate the power and energy consumption of hosts [40].

$$PM_i(t) = k * PM_{i,max} + (1-k) * PM_{i,max} * U_{i,cpu}(t) \quad (2)$$

$$E = \int_{t_0}^{t_1} P_i(t)\, dt \quad (3)$$

Where $PM_{i,max}$is the maximum power consumption of host i, k is the fraction of power consumption when the host i is in idle state and $U_{i,cpu}(t)$ is the CPU resource utilization of the host on time t. E is the energy consumed by host i from start time t_0 to end time t_1.

SLA violation level is measured by two compounds [40]; SLA violation Time per Active Host (SLATAH) and Performance Degradation due to Migrations (PDM). The percentage of the time, during which active hosts experienced the CPU utilization of 100%, called SLA violation Time per Active Host (SLATAH). SLATAH calculates as shown in (4). The overall performance degradation by VMs due to migrations is called Performance Degradation due to Migrations (PDM). PDM is computed as shown in (5). The reasoning behind the SLATAH is the reflection that if a host serving requests are experiencing 100% usage, the performance of the requests is restricted by the host capacity; therefore, VMs are not being provided with the need performance level. SLA violation is defined as shown in (6).

$$SLATAH = \frac{1}{N}\sum_{i=1}^{N}\frac{T_{si}}{T_{ai}} \quad (4)$$

$$PDM = \frac{1}{M}\sum_{j=1}^{M}\frac{C_{dj}}{C_{rj}} \quad (5)$$

$$SLA = SLATAH * PDM \quad (6)$$

Where N is the number of hosts, T_{si} is the total time during which the host experienced the utilization of 100% leading to an SLA violation, T_{ai} is the total of hosts in active mode. M is the number of VMs, $C_{dj}$is the estimate of performance degradation of the VM caused by migration, C_{rj} is the total CPU capacity requested by VM during its lifetime.

6. RESULT& ANALYSIS

In this section, the results of our experiments will be discussed. The following experiments are divided into three classes. In the first class of experiment, the method of prediction workload of the first layer of MLD will be evaluated and discuss. Also, one class support vector machine model in the second layer of MLD will be analyzed. In the second class, the performance of the MLD scheme will be evaluated with a different real workload according to table 4. The real workload is mixed with DDoS attack resource depletion. The attack ratio is 10% of the real workload. The third class of experiment, the MLD will be tested under a various attack ratio. The DDoS attacks ratio are 10%, 20% and 50% of the real workload. For all experiments, the static threshold uses to detect the overutilization hosts. The static threshold for CPU and memory are 0.8290, 0.7651 respectively. Also, all experiments use VM selection policy called the Random Selection (RS) is proposed in [39]. RS was selected to avoid the overhead that another VM selection policies can cause.

6.1. MLD layers Evaluation

Both of MLD layers will be evaluated according to prediction accuracy (R2), the root means square error (RMSE), and the Percentage of Predictions (25) (PRED (25)) metrics for the rigid regression model. The best values for R2 and PRED (25) are close to one and the worst values are close to zero. The best value for RMSE is 0. For the second layer which contains one class support vector machine model, precision (P), recall(R), F-scores, and accuracy.

6.1.1. MLD layer one: Dynamic Threshold for Workload Size

Figure 3 shows the performance of the proposed prediction model over the day. The proposed method tested on Google Cluster Data (GCD) which published in [38]. For figure 3, both predicted values and real values are similar most of the time. At the end of curves, the penalty term success to avoid the over fitting term as the advantage of rigid regression. Google cluster is used in different world sides. the size of people requests jobs form google is varied for people behavior and cultures. Therefore, the increasing or decreasing of people requests can have done suddenly. this is what can be explained the highest difference between the prediction values and real values done at for change the people requests suddenly. The one class support vector machine model in the second layer will be able to avoid this problem.

Table 4 shows the numerical value of the evolution metrics. According to the previous workload size requested and its time, the method will predict the size of the workload request at the current hour. If the dynamic threshold is lower than the real submitted workload, the cloud system will suffer from DDoS attacks. when DDoS attacks start, we recommend increasing the time of jobs spend at schedule stage to avoid job failure at the scheduling stage. The next section will discuss how the DDoS attacks will detect.

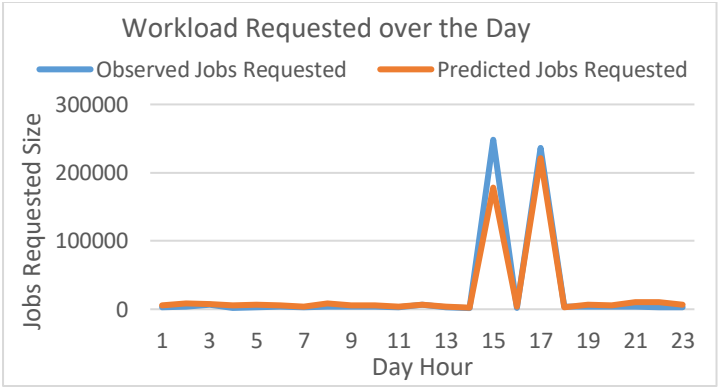


Figure 3. The Performance of the Proposed Prediction Model.

Table 4. The Metric Evolution of the Proposed Prediction Model.

Metric	R^2	RMSE	PRED(25)
Value	0.97433	0.1838	1

6.1.2. Layer Two: One Class Support Vector Machine (OCSVM)

Table 5 shows the result of evolution metrics for CPU and memory OCSVM model. For table 5, the resulting model of CPU and memory results are very similar. The best values for P, R, F score, and accuracy are closed to 1 and the worst values are close to zero. The P-value referees to the ratio of the true positive values to the sum of the true positive values and the false positive values. When the false prediction value minimizes as possible then the P-value becomes high which mean that the OCSVM model is good. The R-value expresses about the ratio of true positive to the sum of true positive and false negative values. When the false negative value minimizes as possible, the R-value becomes high. According to P and R values, the proposed OCSVM model success to reduce the false positive and false negative alarms. The F value is a weighted average of P and R-value. The accuracy measures how the real and predicted values are similar.

Table 5. The Evaluation Performance for CPU and Memory OCSVM Model.

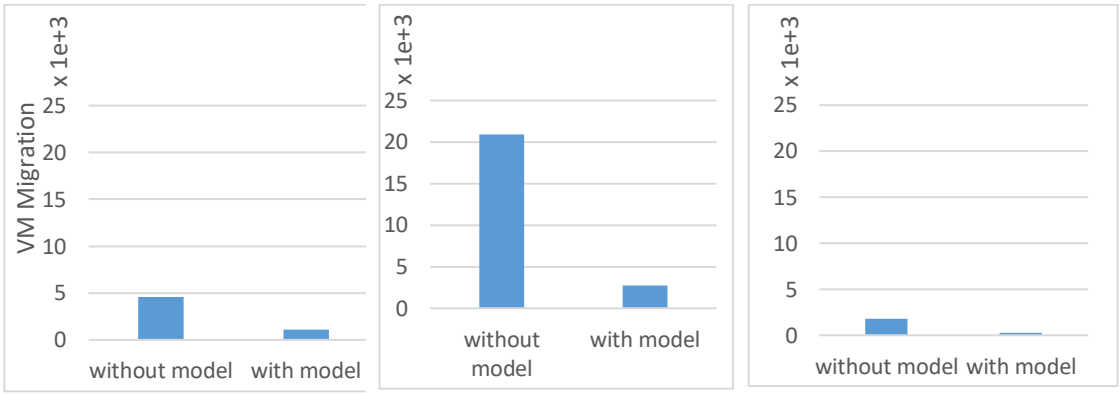
CPU				Memory			
P	R	F	Accuracy	P	R	F	Accuracy
1	0.997	0.998	0.997	1	0.998	0.999	0.998

6.2. MLD Performance under DDoS Attack

The comparison between cloud computing performance with and without the MLD scheme will discuss. The MLD will test with three different real workloads mixed with DDoS attacks called D1, D2, and D3. The MLD will evaluate according to the number of VM migration, energy consumption, the overall SLA violation, and the number of hot hosts.

6.2.1. The Number of VM Migration

According to table 5, three data set have a different size of workload requested. The largest data set is D1 and the smallest is D2. Figures 4a, 4b, and 4c display the effect of the proposed MLD scheme on the number of VM migration for various data set. It clears from figure 4 that the number of VM migration is produced on the smallest data set D2 is the highest than the number of VM migration of others datasets. The main explanation for this issue that the cloud environment has a lot of empty allocation able to receive which VM will migrate. Also, the majority of VM in D2 is from micro VM which can migrate simply. The MLD has a good effect to detect and mitigate the DDoS attacks. The MLD reduced the number of VM migration by 76.3%, 86.7%, and 84.9% for D1, D2, and D3 respectively. The huge VM migration is produced from that DDoS attacks consume more resources than are have. So, the VM migrated from host to another to find resources are required. When the MLD achieves success to detect and remove the DDoS attacks, the VMs are caused the DDoS attacks remove, then the number of VM migration decrease.



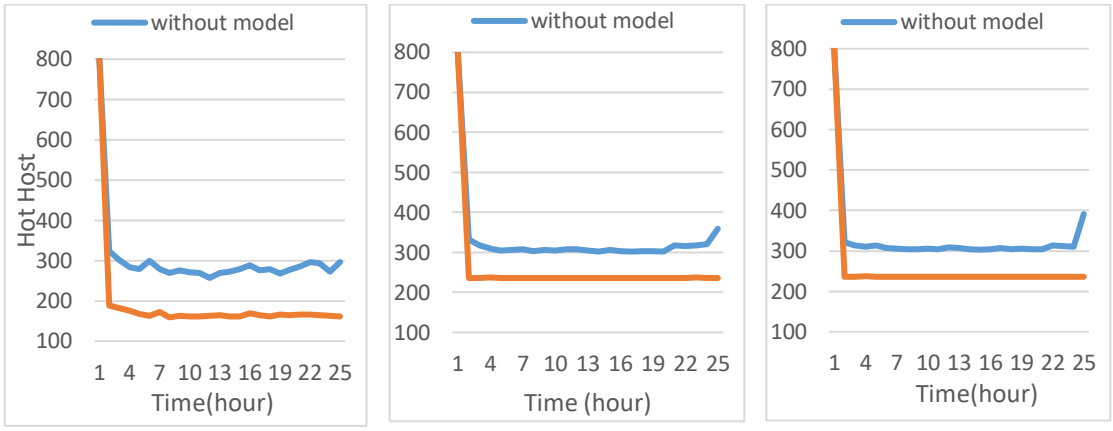
(a) (b) (c)

Figure. 4. The Performance of The MLD for The Number of VM migration with various datasets: dataset D1; (b) Dataset D2; (c) Dataset D3.

6.2.2. The Number Hot Host

When the number of VM migration is increasing, thehothost's number is going to increase as well. The main reason is that the VM of the DDoS attacks is looking for free provisions for hosting. If there are no free allowances, an inactive host wakes up. Therefore, the hothost's number is increasing. Figures 5a, 5b, and 5c demonstrate the effect of the proposed MLD scheme on the number of hot host for various data set. It clears that the MLD achieves success to reduce the number of the hot host. In addition, the behavior of the hot host under the MLD is more stable than the hot host without the MLD. For figures 5, the number of the hot host will increase at last hour which it is supporting that the cloud system is going to fail in the near future. In the average, the MLD reduced the number of the hot host by 40.9%, 24%, and 24% for D1, D2, and D3respectively

.

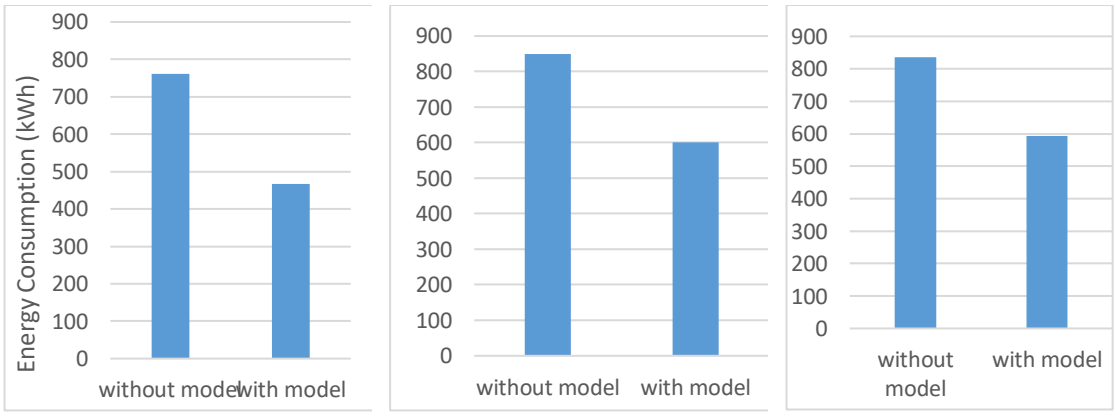


(a) (b) (c)

Figure 5. The Performance of The MLD for The Number of Hot Host with various datasets: Dataset D1; (b) Dataset D2; (c) Dataset D3.

6.2.3. Energy Consumption

Both the numbers of VM migration and the hot hosts' number have a high impact on energy performance. According to equations 2 and 3, the change rate of power is effective in energy consumption. The small number of VM migration due to the rate change of power and energy consumption becomes low. Also, the small number of hot host aims to reduce the amount of energy consumption. Therefore, energy consumption becomes low. Figures 6a, 6b, and 6c explain the effect of the proposed MLD scheme on the energy consumption for various data set. For figure 6, it notes that the dataset D2 is the highest in energy consumption because it has the highest number of VM migration and the average highest number of the hot host. The MLD achieves success to reduce energy consumption by 38.6%, 29.1%, and 29% for D1, D2, and D3 respectively



(a) (b) (c)

.

Figure. 6. The Performance of The MLD for The energy consumption with various datasets: Dataset D1; (b) Dataset D2; (c) Dataset D3.

6.2.4. Overall SLA Violation

Two main parameters are affected in the SLA violation are; the degradation due to VM migration, and the time of full utilization host according to equation 6. The degradation due to VM migration is calculated by the memory of VM migrated, the network bandwidth, and the number of VM migration. Where the network bandwidth in cloud computing become more developed, then the most effective in the degradation due to VM migration is the memory of VM migrated and the number of VM migration. The number of VM migration defines as the amount of VM migrated from host to reduce the overutilization of host and the amount of VM migrated from host to become inactive. Figures 7a, 7b, and 7c appear the effect of the proposed MLD scheme on the overall SLA violation for various data set. According to table 5, the dataset D1 is the largest than the dataset D2 and D3. The overall SLA violation of D1is the largest because the number of extra and high VMs in D1 is higher than extra and high VMs in D2 and D3. The extra and high VMs have the highest memory than micro and small VMs. Therefore, the dataset D1 has a higher SLA violation than the others. Both D2 and D2 have similar SLA violation because they have a similar total number of extra and high VMs. However, the dataset D3 has a huge number of micro VMs, the SLA violation of D3 is lower than the SLA violation of D1. Consequently, the huge number of micro VMs is less effective in degradation due to VM migration and the overall SLA violation. The MLD achieves success to reduce the overall SLA violation by 99.73%, 98.8%, and 97.6% for D1, D2, and D3 respectively.

6.3. DDoS Attacks Ratio

In this section, the stability and reactivity of MLD will evaluate under a various DDoS attacks ratio. The MLD will be tested under 10%, 20%, and 50% DDoS attack ratio respectively. Figures 8a, 8b, and 8c show the effect of a various DDoS attacks ratio on the proposed MLD scheme on the number of VM migration, the energy consumption, and the overall SLA violation for the dataset D1. Figure 8 demonstrates that the MLD has the ability to maintain good performance.

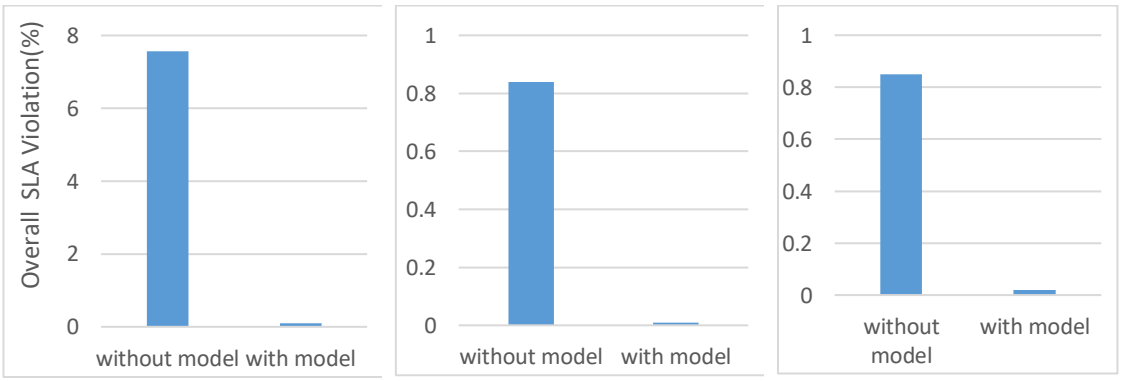
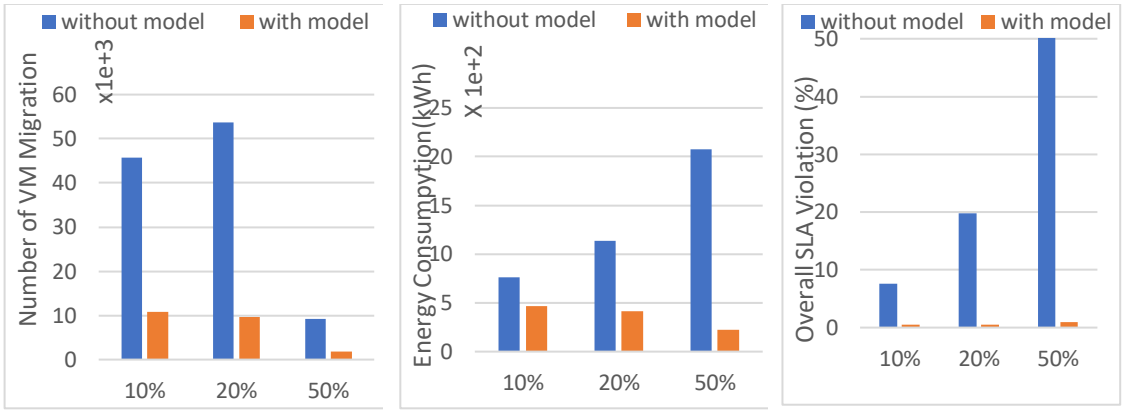


Figure 7. The Performance of The MLD for The Overall SLA Violation with various datasets: Dataset D1; (b) Dataset D2; (c) Dataset D3.

under a various attack ratio. For the number of VM migration, it clears that the number of VM migration is increasing with increase the attack ratio in the case of the cloud system without the MLD scheme. At 10% and 20% attack ratio, the number of VM migration increases with increasing attack ratio. While at 50% attack ratio, the number of VM migration decrease due to

the high attack ratio has a lot of the DDoS attacks machine. Also, with high attack ratio, DDoS attack VM consumes all available resource in its host which due to that no host has free resources and cannot receive any VMs. With the MLD scheme, the number of VM migration reduces.



(a) (b) (c)

Figure 8. The Performance of The MLD under several Attack ratios for the dataset D1: The number of VM Migration; (b) Energy Consumption; (c) Overall SLA Violation.

For energy consumption, the MLD achieves success to reduce energy consumption at a different attacks ratio. It observes that the energy consumption of the cloud system without the MLD increases with increasing attacks ratio. Otherwise, the MLD reduces energy consumption with increasing attack ratio. The main explanation is that the MLD detects and removes the DDoS attacks and only the good VMs are keeping run. In addition, the high attack ratio is equivalent to the number of good VMs are less. Therefore, the energy consumption with the MLD scheme becomes less at high attack ratio. For the overall SLA violation, it notes that the MLD still stable with a different attack ratio. Also, MLD is able to reduce the SLA violation at all attack ratios. In the case without MLD, with increasing attack ratio, the overall SLA violation is growing because the number of attack machines is higher than good machines. It observes that the MLD scheme detects and removes the DDoS at early times for all attacks ratio, for that, the number of hosts arrives to lower utilization early.

7. CONCLUSION AND FUTURE WORK

This paper discussed the DDoS attack detection and mitigation problem. This paper focused on the DDoS attack resource depletion category. The MLD scheme is proposed to detect and mitigate the DDoS attacks. The MLD scheme consists of two layers. The first layer contains an alarming system, the main objective of alarm system sends alerts to the cloud system management when the DDoS attacks start. The alarm system is based on predicting the size of the workload requested. The predicted workload is used as a dynamic threshold to compare with real workload requested at the current time. According to the result of comparison between the dynamic threshold and real workload, the alarm system sends its alerts or not. When the DDoS attack starts, we recommend that extend the time of schedule stage to avoid VM failure in the scheduling stage according to increasing the size of requested jobs than normal. The second layer includes an anomaly detection system based on one class SVM. The main benefit of the one class SVM is that is more robust and less sensitivity by outliners. Also, a labeled data for training does not required.

The MLDwas tested through a variety of real workload of different size mixed with DDoS attacks. The number of VM migration, the number of hot hosts, energy consumption, and the overall SLA violations have been used to evaluate the performance of cloud computing under the MLD scheme. The MLD scheme provides great help in detecting and mitigating DDoS attacks. The results show that the MLD scheme reduces the number of VM migration, the number of hot hosts, the energy consumption, and the overall SLA for a various real workload that were used in the evaluation process. In addition, the MLD scheme tested with various DDoS attacks ratio 10%,20%, and 50%. The MLD showed more stability and reactivity for all tested DDoS attack ratio. In the future, our future plan aims to add a new layer to the MLD scheme. The proposed new layer will containa prevention method. The anticipated new layer will elevate the MLD scheme to a complete defense system. Also, in the future, we aim to test our proposed scheme in a real cloud system.

REFERENCES

[1] Linlin Wu, Saurabh Kumar Garg, Steve Versteeg, and Rajkumar Buyya, (2014) "SLA-Based Resource Provisioning for Hosted Software-as-a-Service Applications in Cloud Computing Environments", IEEE Transactions On Services Computing, Vol. 7, No. 3, pp. 465-48.

[2] Per Wikström, Lucien A. Terens, and Heinz Kobi, (2000) "Reliability, Availability, and Maintainability of High-Power Variable-Speed Drive Systems", IEEE Transactions On Industry Applications, Vol. 36, No. 1, pp. 231-241.

[3] Saurabh Singh, Young SikJeong, Jong HyukPark, (2016) "A survey on cloud computing security: Issues, threats, and solutions", Journal of Network and Computer Applications, vol.75, pp. 200-222.

[4] Choi, Junho, Choi, Chang, Ko, Byeongkyu, Choi, Dongjin, Kim, Pankoo, (2016) "Detecting web basedDDoS attack using MapReduce operations in cloud computing environment", Journal of Internet Services and Information Security, Vol 3, Issue 3-4, pp. 28-37.

[5] Raneel Kumar, Sunil Pranit Lal, AlokSharam, (2016) "Detecting Denial of Service Attacks in the Cloud", 2016 IEEE 14th Intl Conf. on Dependable, Autonomic and Secure Computing, 14th Intl Conf. on Pervasive Intelligence and Computing, 2nd Intl Conf. on Big Data Intelligence and Computing and Cyber Science and Technology Congress, pp.309-316.

[6] Gaurav Somani, Manoj Singh Gaur, DheerajSanghi, Mauro Conti, Rajkumar Buyya, (2017) "DDoS attacks in cloud computing: Issues, taxonomy, and future directions", Computer Communications, vol. 107, pp. 30-48.

[7] Arbor Networks (2015), Worldwide infrastructure security report volume XI.

[8] Kaspersky Labs, Global IT security risks survey 2014 (2014) - distributed denial of service (DDoS) attacks.

[9] Rashmi V. Deshmukh, Kailas K. Devadkar, (2015) "Understanding DDoS Attack & Its Effect In Cloud Environment", Procedia Computer Science, Vol. 49, pp.202-210.

[10] OpeyemiOsanaiye, Kim-Kwang Raymond Choo, MqheleDlodlo, (2016)" Distributed denial of service (DDoS) resilience in cloud: Review and conceptual cloud DDoS mitigation framework", Journal of Network and Computer Applications, vol. 67, pp.147-165.

[11] Tag Man, Just one second delay in page-load can c 7% loss in customer conversions, 2013, (http://www.tagman.com/mdp- blog/2012/03/just- one- second- delay- in- page- load- can- cause- 7- loss- in- customer- conversions/).

[12] G. Somani, M.S. Gaur, D. Sanghi, M. Conti, (2016) "DDoS attacks in cloud computing: collateral damage to non-targets", Computer Networks, Vol. 109, No. 2, pp. 157-171.

[13] Ruchi Mehta, (2017)" Distributed Denial of service Attacks on Cloud Environment", International Journal of Advanced Research in Computer Science, Volume 8, No. 5, pp. 2204-2206.

[14] Jingqi Yang, Chuanchang Liu, Yanlei Shang, Bo Cheng, Zexiang Mao, Chunhong Liu, LishaNiu, Junliang Chen,(2014) "A cost-aware auto-scaling approach using the workload prediction in service clouds",Information Systems Frontiers, Vol. 16, No. 1, pp. 7-18.

[15] Ahamed Radhwan, Mahmoud Kamel, Mohamed Y.Dahab, Aboul Ella Hassanien, (2015)" Forecasting Exchange Rates: A Chaos-Based Regression Approach", International Journal of Rough Sets and Data Analysis, Vol. 2, No.2, pp. 38-57.

[16] Wei Fang, ZhiHui Lu, Jie Wu, ZhenYin Cao, (2012)"RPPS: A Novel Resource Prediction and Provisioning Scheme in Cloud Data Center",2012 IEEE Ninth International Conference on Services Computing, pp. 609-616.

[17] Chunhong Liu, Yanlei Shang, Li Duan, Shiping Chen, Chuanchang Liu, Junliang Chen, (2015) "Optimizing Workload Category for Adaptive Workload Prediction in Service Clouds",International Conference on Service-Oriented Computing, pp. 87-104.

[18] Chunhong Liu, Chuanchang Liu, Yanlei Shang, Shiping Chen, Bo Cheng, Junliang Chen, (2017)" An adaptive prediction approach based on workload pattern discrimination in the cloud", Journal of Network and Computer Applications vol. 80, pp. 35-44.

[19] Yazhou Hu, Bo Deng, Fuyang Peng, Dongxia Wang, (2016)" Workload Prediction for Cloud Computing Elasticity Mechanism", 2016 IEEE International Conference on Cloud Computing and Big Data Analysis, pp. 244-249.

[20] Rodrigo N. Calheiros, EnayatMasoumi, Rajiv Ranjan, RajkumarBuyya, (2015) "Workload Prediction Using ARIMA Model and Its Impact on Cloud Applications' QoS", IEEE Transactions On Cloud Computing, VOL. 3, NO. 4, pp. 449-458.

[21] Gopal KirshnaShyama, SunilkumarS.Manvi, (2016)" Virtual resource prediction in cloud environment: A Bayesian approach", Journal of Network and Computer Applications, Vol.65, pp.144-154.

[22] Mohammed H. Sqalli, Fahd Al-Haidari, Khaled Salah, (2011) "Enhanced EDoS-Shield for Mitigating EDoS Attacks Originating from Spoofed IP Addresses", 2011 Fourth IEEE International Conference on Utility and Cloud Computing, Victoria, NSW, pp.49-57.

[23] Fahd Al-Haidari, Mohammed H. Sqalli, Khaled Salah, (2012)"Enhanced EDoS-Shield for Mitigating EDoS Attacks Originating from Spoofed IP Addresses", 2012 IEEE 11th International Conference on Trust, Security and Privacy in Computing and Communications, pp.1167-1174.

[24] B. Saini, G. Somani, (2014) "Index page based EDoS attacks in infrastructure cloud", International Conference on Security in Computer Networks and Distributed Systems, Springer, pp.382-395.

[25] Siqin Zhao, Kang Chen, Weimin Zheng, (2009)" Defend Against Denial of Service Attack with VMM", 2009 Eighth International Conference on Grid and Cooperative Computing, pp. 91-96.

[26] Shui Yu, Yonghong Tian, Song Guo, Dapeng Oliver Wu, (2014) "Can We Beat DDoS Attacks in Clouds", IEEE Transactions On Parallel And Distributed Systems, Vol. 25, No. 9, pp. 2245-2254.

[27] Min Du, Feifei Li, (2015) "ATOM: Automated Tracking, Orchestration and Monitoring of Resource Usage in Infrastructure as a Service Systems", 2015 IEEE International Conference on Big Data (Big Data), pp.271-278.

[28] G. Somani, A. Johri, M. Taneja, U. Pyne, M.S. Gaur, D. Sanghi, (2015)" DARAC: DDoS mitigation using DDoS aware resource allocation in cloud", 11th International Conference, ICISS, pp. 263–282.

[29] PengXiao, WenyuQu, HengQi, ZhiyangLi, (2015)" Detecting DDoS attacks against data center with correlation analysis", Computer Communications Vol. 67, pp.66-74.

[30] Zhiyuan Tan, Aruna Jamdagni, Xiangjian He, Priyadarsi Nanda, Ren Ping Liu, (2014)" A System for Denial-of-Service Attack Detection Based on Multivariate Correlation Analysis" IEEE Transactions On Parallel And Distributed Systems, Vol. 25, No. 2, pp.447-456.

[31] MouhammdAlkasassbeh, Ahamed B.A Hassanant, Ghazi Al-Naymat, Mohammad Almseidin, (2016) "Detecting Distributed Denial of Service Attacks Using Data Mining Techniques", International Journal of Advanced Computer Science and Applications, Vol.7, No.1, pp.436-445.

[32] Gupta S, KumarP., (2013)" Vm Profile Based Optimized Network Attack Pattern Detection Scheme for DDoS Attacks in Cloud", International Symposium of security in computing and communications (SSCC2013), pp.255-261.

[33] A. M. Riad, Ibrahim Elhenawy, Ahmed Hassan, Nancy Awadallah, (2013), International Journal of Computer Networks & Communications (IJCNC), Vol.5, No.5, pp.195-208.

[34] Markus Goldstein, Seiichi Uchida, (2016)" A Comparative Evaluation of Unsupervised Anomaly Detection Algorithms for Multivariate Data",PLoS ONE, Vol.11, No.4, pp.1-31.

[35] MehiarDabbagh, BechirHamdaoui, Mohsen Guizani, Ammar Rayes, (2015)" Energy-Efficient Resource Allocation and Provisioning Framework for Cloud Data Centers", IEEE Transactions On Network And Service Management, Vol. 12, No. 3, pp.337-391.
[36] Amazon EC2, (2015) http://aws.amazon.com/ec2/.
[37] Sheng Di, Derrick Kondo, WalfredoCirne, (2012)" Characterization and Comparison of Cloud versus Grid Workloads", IEEE International Conference on Cluster Computing, pp. 230-238.
[38] Traces of Google Workloads, (2015) http://code.google.com/p/ google cluster data/.
[39] Nguyen TrungHieu, Mario Di Francesco, Antti Yl ¨a-J¨a¨ aski, (2017) "Virtual Machine Consolidation with Multiple Usage Prediction for Energy-Efficient Cloud Data Centers", IEEE Transaction On Services Computing, Vol. Pp, No. 99, Pp. 1-14.
[40] A. Beloglazov and R. Buyya, (2012) "Optimal online deterministic algorithms and adaptive heuristics for energy and performance efficient Dynamic consolidation of virtual machines in cloud data centers," Concurrency and Computation: Practice and Experience, pp. 1397–1420.