OJST

Open Journal of Science and Technology
http://readersinsight.net/OJST

Research Article

DYNAMIC RESOURCE ALLOCATION IN CLOUD COMPUTING BASED ON SOFTWARE-DEFINED NETWORKING FRAMEWORK

Arwa Mohamed[1]*, Mosab Hamdan[2], Ahmed Abdelazizb[3], Sharief F. Babiker[4]

[1]Faulty of Engineering, University of Khartoum, khartoum, 11115, Sudan & Faculty of Engineering, University of Science and Technology, khartoum,11115, Sudan.
[2]School of Electrical Engineering, Faculty of Engineering, Universiti Teknologi Malaysia, Johor,81310, Malaysia.
[3]Faculty of Computer Science, Future University, Khartoum ,11115, Sudan.
[4]Faculty of Engineering, University of Science and Technology, khartoum,11115, Sudan.

***Corresponding Author email:** arwajewels@gmail.com

ABSTRACT

cloud computing has become more powerful with the inclusion of software-defined networking (SDN) in its environment. In Cloud Data Centers (CDCs), an important research issue is how to forecast and allocate resources efficiently whilst achieving Quality of Service (QoS) of users request with minimal overall power consumption; taking into account the frequent changes in resource requirements. In this paper, we propose a Supervisor Controller-based Software-Defined Cloud Data Center (SC-boSD-CDC) framework for dynamic resource allocation and prediction of cloud computing-based SDN. In this proposed module, Genetic Algorithm (GA) is proposed to deal with the multi-objective problem of dynamically forecasting the utilization of resources in both compute nodes and links bandwidth of network as well as energy consumption in the Cloud Data Center (CDC). Furthermore, a Virtual Machines (VMs) placement algorithm is also proposed to allocate computing resources and routing algorithms to choose the proper bandwidth links between switches; resulting in increased CPU and memory utilization and reduction in overall power consumption.

Article History

Submission: July 19, 2020
Revised: August 30, 2020
Accepted: September 21, 2020

Keywords:
Software-Defined Networking; Cloud Computing; Genetic Algorithm; Resource Allocation; Data Center

Abbreviations: NIL

doi
10.31580/ojst.v3i3.1668

INTRODUCTION

Cloud Computing (CC) (1) is a technology that facilitate remote access and manipulation of data and applications through the internet. CC provides the possibility of pay per use of the services offered by the cloud provider without need for application installation. Also, it allows powerful high performance computing through the use of centralized storage, memory, process and bandwidth. As stated by the National Institute of Standards (NIST), various types of service models provided by cloud computing include Infrastructure as a Service (IaaS), Software as a Service (SaaS) and Platform as a Service (PaaS). Additionally, NIST also showed the various models of rendering cloud services to cloud clients as private cloud, public cloud, and hybrid cloud (2).

CC built upon data center infrastructures is designed to apply virtualization technology. Virtualization technology is a mechanism that enables the virtualization of data center resources such as CPU, memory, storage unit switches, and links. Using hypervisors in the data centers allows the deployment of Virtual Machines (VMs), which are shared and rented out to cloud users (tenants). Some of the features provided by this technology include the host's flexibility in configuring many VMs on it which consequently increases the utilization and flexibility of computing resources of a network. It also facilitates the allocation of the resources that can be changed dynamically, depending on client requests. However, the inflexibility in the management of virtual resources becomes apparent when the virtualization technology is applied in CC.

Software-Defined Networking (SDN) (3) is a technology that restructures the old network architecture by separating the control plane from the forwarding plane. This separation, as well as the centralized controller, makes the network more powerful adaptable to sudden changes. An OpenFlow protocol (4) allows connection between switches and the SDN controller. SDN has many features that has great benefits when adopted in cloud computing. Some these benefits include; dynamically adapting to changing workload, enhanced network security, programmable control plane, virtualized network, global view of data center network, consolidation of Virtual Machine Monitor (VMM) with the cloud computing management (5). SDN facilitates the process of integrating a virtual environment into the physical environment provided by network providers to solve the problem of management in the traditional CDC.

Operational costs of the data centers are affected by energy and power consumption. Notably, the energy consumed in cloud data centers has considerably increased in recent times. Hourly energy consumption in 2020 is expected to reach 140 billion kilowatts, resulting in an annual spending of $13 billion (6). Currently, there is no general approach or algorithm that optimizes computing resources with network resources and activities. Combination of computing and network resources optimization will be exceptional due to the lack of technologies that control them together (7). In the cloud data center, most researchers focus on allocating resources either in compute nodes or link bandwidth separately within the network while ignoring the optimization of resource utilization and power consumption (8, 9).

In this paper, we propose a Supervisor Controller based on Software-Defined Cloud Data Center (SC-boSD-CDC) to jointly formulate the problem of resource prediction and allocation in the CDC in relation to computing and network resources. OpenStack and SDN technologies are used to achieve the aim of this work. SDN is used to achieve dynamic control of network configuration like network monitoring, and bandwidth allocation with the help of computing resource management techniques. We also proposed a Multi-Objective Genetic Algorithm (MOGA) to predict resource utilization in compute nodes and bandwidth link utilization between switches. The VM placement algorithm is subsequently used to allocate computing resources while routing algorithms are used to choose the proper bandwidth links between switches; thereby leading to increased CPU and memory utilization and optimal overall power consumption.

This paper is organized into a group of parts as follows; A review of pertinent works related to allocation in CDC based on SDN are presented in Section II. The proposed system framework components are illustrated in Section III while the conclusion and future direction is presented in Section IV.

RELATED WORK

RESOURCE ALLOCATION TECHNIQUES IN CC.

To fulfill the QoS for any client request while also taking into account the overall power consumption in the data centers, researchers have used two methods for resource allocation in CDCs, namely: 1) predicting utilization of resources and then allocating it. 2) Only resources allocation. As shown in Fig. 1. The research work carried out in (10) represents optimal task schedules by using skewness algorithm to minimize total time while load balancing in VMs migration is used to reduce energy. This is a two-stage implementation; the need to predict resources, followed by the process of virtual machine migration. Hence, three thresholds were identified: the hot and cold spot of each machine is identified by predicting the resource demands and resource allocation of all virtual machines .Whereas the Hot spot represents the overloading of the server, the cold spot shows the utilization value below the threshold. The warm spot illustrates the level of utilization that is sufficient for running the server. Note that each type of resource can have different thresholds at the time of allocation. Depending on the temperature of the server, the VM must migrate to other servers while ensuring that the destination server must not become a hot spot. In contrast, choosing this destination server is based on the reduction of the skewness value; it represents the second stage of implementation, which explains the irregularity in the measurement of resource usage in the server. The skewness algorithm achieves overload avoidance with less time. Thus, in the context of huge data center, VM migration becomes a robust process. Researchers have proposed approaches to address the problem encountered by using multi-agent-based migration. Tseng *et al.* (11) proposed a MOGA to predict the utilization of resources and energy consumption in the CDC dynamically. To resolve the problem of resource allocation, a GA is proposed to forecast the resource requirement of the next time slot, depending on collected historical data from previous slots for both VMs and PMs. The prediction result was compared with the grey forecasting model and it was noted that the proposed model increased the average utilization of CPU and memory with decreased energy consumption of the cloud data center to satisfy the QoS with respect to the overall energy consumption for each client request.

Alkayal *et al.* (8) proposed allocation in cloud computing environments with multi-objective optimization (MOO) approach by exploiting a particle swarm optimization (PSO) algorithm. The process of allocation was done in three stages as follows: (i) The clients and providers negotiate to choose the Data Center (DC), (ii) Scheduling of tasks in DCs, and (iii) allocation of VMs to PMs. Thus, by reducing negotiation time and mapping time, both standby time and productivity are consequently improved. Moreover, when using VM migration technology, the resource utilization efficiency was increased by loading the balance. Additionally, power consumption was reduced by shutting down unused hosts. This results in an improvement in overall productivity.

Resource-efficient and performance-aware cloud infrastructures which focuses on the energy-efficient optimization of PMs to reduce the operational cost was presented in (9). The authors used a dynamic VM consolidation approach to achieve this. Column generation technique (12) was used to decrease consumption of energy in servers and network resources in the DC. This joint optimization of power consumption was achieved in two stages. The first one being optimization of power in server and/or network while the second stage deals with the migration cost of VMs, which is subjected to heterogeneity in resource and bandwidth constraints for both workload and server in the cloud computing center. The authors compared the proposed column generation technique with integer quadratic program IQP and a 2% increase in the performance accuracy of CG was reported over IQP. The reported numerical results show better optimization at the data center regarding power saving compared to VM placement heuristic algorithms.

A heuristic algorithm called a non-dominated sorting genetic algorithm was used to allocate resources in the cloud data center for both computational and network resources a presented in (13). The authors sought to strike balance between maximizing the time to complete tasks and reducing the power consumption of servers and network switches. This approach applied a Meta framework on homogeneous hosts in the same DC.

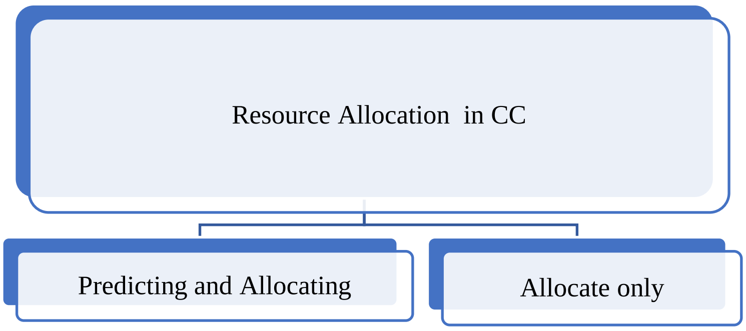

Fig. 1. Classifications of Resource Allocation Techniques.

B. CC BASED ON SDN

SDN technology-based on cloud infrastructure has rarely been studied in the literature for resource prediction and allocation. For example, in (14), it is used to predict the resources required from each application in SDN-based CDCs. The research work carried out in (15) states that the networking resources in Virtual Infrastructures (VIs) are difficult to manage. Therefore, they formulated an online allocation for virtual infrastructure as an optimal MIP while considering all traditional aspects and SDN challenges that appear when applied to resource allocation at the CDC. Moreover, they presented a mechanism for getting a linear program in addition to a heuristic rounding by relaxing the MIP. Two principal mechanisms were compared with QVIA-SDN, resulting in higher internal latency. Furthermore, QVIA-SDN, SDN-aware allocation is used to minimize the data center usage and improve the QoS understanding by hosted tenants. Son *et al.* (7) presented a mechanism for the fulfillment of QoS and power consumption in SDN enabled CDCs by combining the allocation of compute and network resources. This method focuses on the techniques to achieve a balance between energy efficiency, QoS, and service level agreement (SLA), which allows allocation of VMs and networks on PMs and switches.

(16) considered link bandwidth and flow table as essential resources in SDN. Hence, they proposed the joint allocation of link bandwidth and flow table in SDN to meet different types of application controls. Moreover, Elastic Tree was used as a power optimizer to dynamically adjust the data center network power consumption by monitoring traffic patterns. Thus, the optimal network path is recomputed continuously by the optimizer and more capacity is brought online when traffic increases; this process continues until it reaches the full network capacity. On the other hand, switches and links are turned off when traffic decreases. Elastic Tree was used to achieve tradeoffs between performance, robustness, and energy (17).

Another approach, presented in (18), routes packets from source to destination host by finding the best energy-efficient path. It is implemented using a modified Swarm intelligence meta-heuristic algorithm called Particle Swarm Optimization (PSO). The fitness value of PSO is calculated based on the energy of node connectors of each switch, which represents the optimal selection path from source to the destination in switch level by level. The algorithm was developed on the OpenDaylight controller (19) using the Mininet simulation (20). However, that work did not consider the network requirements when applied in the emulator environment.

PROPOSED SYSTEM

The design goal of our system is to formulate an optimization problem for both prediction and resource allocation in CDC based on SDN. This has several benefits which include maximizing utilization of CPU and memory in VMs and physical machines (PMs) within computing nodes and maximizing utilization of link bandwidth in network switches while minimizing power consumption in the data center. Hence, we proposed our designed framework on top of cloud platforms and SDN controller software. Fat-tree topology is used to construct the architecture of the system.

Cloud Management Platform Module(CMPM):

CMPM is an OpenStack emulator that is responsible for computing resources. It uses Nova-compute to manage compute nodes, while Neutron is responsible for managing the virtual networks for VMs. Openstack ceilometer component gathers computing resource (e.g. CPU utilization and RAM usage) measurements from each compute node for VMs and PMs. Thereafter, a multi-objective genetic algorithm uses the collected information to predict the utilization of compute nodes.

Networking Module (NM):

SDN controller controls OpenFlow switches that connect compute nodes. It is also responsible for VM to VM bandwidth allocation. SDN controllers monitor flow of data from switch through sFlow agent that is installed in each switch and provide the measurements via a Rest API. The gathered data is exploited by a prediction module to schedule the link bandwidth.

Supervisor Controller based on Software-Defined Cloud Data Center(SC-boSD-CDC)

When requests are received from users, the Supervisor Controller unit decides which VMs are allocated to PMs, and which link is used to carry the VM flow. It also jointly controls the CMPM and the SDN Controller. Decisions are made using the allocation algorithm in addition to the algorithm dedicated to scheduling the network flow. Meanwhile, the algorithmic and scheduling processes are the responsibility of the prediction unit that collects historical data from the cloud and NM. Consequently, Supervisor Controller utilizes the CMPM to initiate VMs on the selected servers as well as make flow up to date with the help of the network manager.

Supervisor Controller can migrate under-utilized VMs and flow, depending on VM migration and bandwidth scheduling algorithms respectively, to a small number of hosts and shutdown idle resources for both PMs and switches to reduce overall power consumption. Fig. .2 describes the proposed system architecture.

SC-boSD-CDC is executed on the top of these software platforms to manage the resources of computing and networking together. The goal of this framework is to merge the CMM and SDN controller to:

- Monitor networking and compute nodes devices in DC for their availability, capacity, and utilization.
- Predict and forecast resource utilization to allocate it based on monitor data.
- Dynamically allocate resources to both networking and compute nodes to maximizing resource utilization and reduce power consumption.

A complete explanation of each component in SC-boSD-CDC is illustrated in the following subsections also shows in Fig. 4.

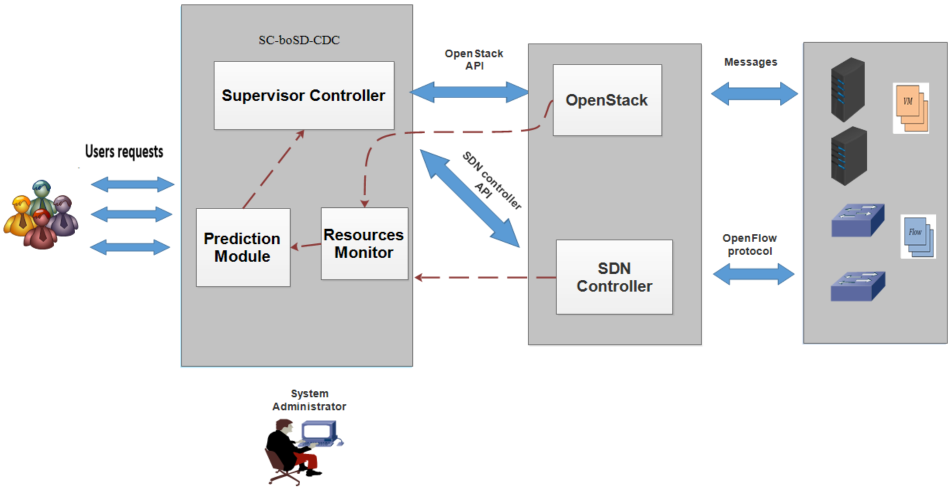

Fig. 2. The System Architecture of the Proposed System

Prediction Module:

Resource allocation for data centers is a Multi-Objective Optimization Problem (MOOP) as illustrated in equation (1).

$$H(y) = \begin{cases} H_1(y) = H_{cpu}(y) \ = \max(cpu - avg) \\ H_2(y) = H_{memory}(y) \ = \max(M - avg) \\ H_3(y) = H_{bandwidth}(y) \ = \max(BW - avg) \\ H_2(y) = H_{energy}(y) \ = \max(E - avg) \end{cases} \tag{1}$$

Where $cpu - avg$ is the average utilization of CPU for n PMs, $M - avg$ is the average value of memory utilization in n PMs, $BW - avg$ represents the bandwidth of link of network and n PMs and are $E - avg$ indicates total consumed energy.

We propose a GA to solve this MOOP of the data center. The resource utilization and power consumption which are predicted dynamically in the CDC using the historical data collected from the cloud and network monitor are also used to reallocate the resources. Notably, GA is faster and more effective in terms of the solution area and search time. The GA outperformed the heuristic algorithm, especially after a solution is obtained from the complete evolution process.

Fig. 3 illustrates the proposed GA flowchart of the resource prediction. A chromosome in GA contains m subparts, which includes a collection of VMs in PM. Each subpart specifies the use of VM resources like CPU usage, memory usage, link bandwidth, and power consumption. Depending on the cloud and network monitoring results, the fitness value of each chromosome can be calculated. Survival of the fittest is represented by the fitness function of the chromosomes. The surviving chromosomes are collected in a mating pool where crossover operation is applied to produce the offspring.

Superior prediction accuracy for CPU usage, memory usage, and link bandwidth and power consumption is achieved at the minimum fitness value of a chromosome. However, a more accurate resource forecasting result can be acquired. Furthermore, to avoid falling in local optimal solutions, the mutation operation must be used. The

GA keeps up the development operation until it attains the VM placement time. Consequently, the GA finishes the development process when the time access for VM placement is reached despite the fact that historical data collection is uninterrupted. Fig. 4. shows the schematic data flow of components of SC-boSD-CDC.

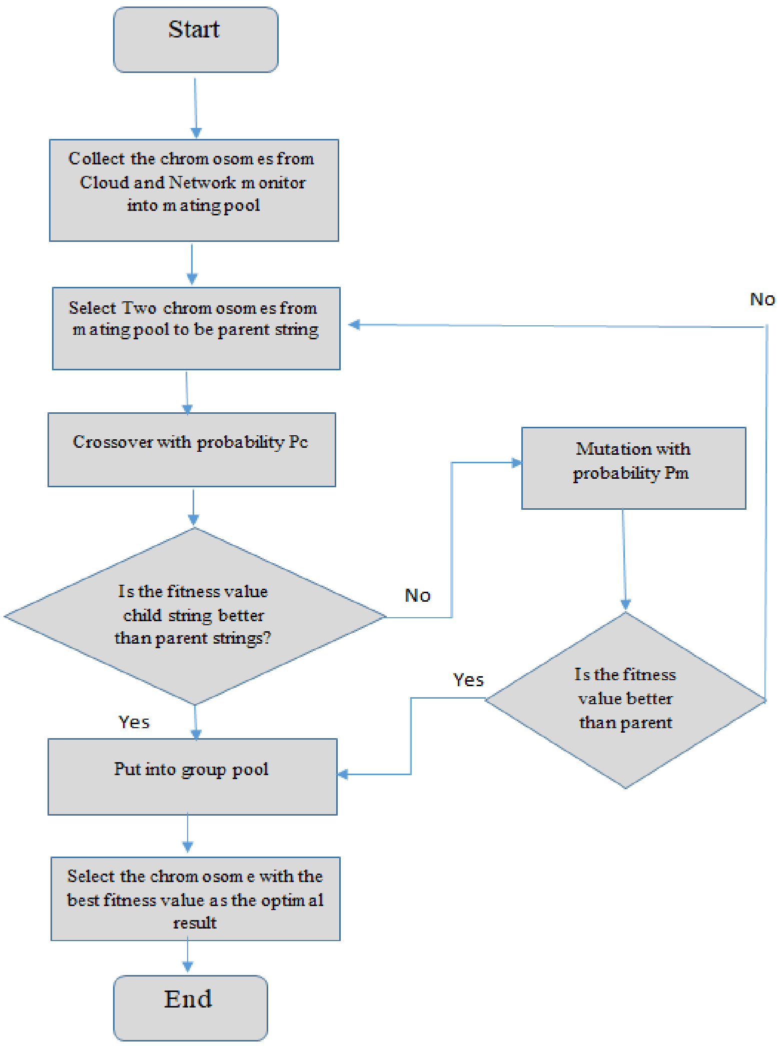

Fig. 3. The Flowchart of GA for Resource Prediction

Cloud Manager Module(CMM):

Cloud Manager controls OpenStack Controller through OpenStack SDK. Furthermore by applying the VMs placement algorithm for allocating CPU and memory resources to maximize utilization and minimize power consumption depending on the prediction model.

Network Manager Module (NMM):

The management of the SDN Controller is performed by the NMM, which is responsible for loading OpenFlow rule tables, QoS configurations, and setting the default path for multi-path load balancing. In addition, it is responsible for the network flow scheduling if an insufficient default path is specified by Network Manager.

Network Monitor(NM):

Network Monitor is responsible for collecting real-time network status from switches and compute nodes. The measurements for each link (bandwidth use and traffic flows) are aggregated using the REST APIs supplied by sFlow-RT.

Cloud Monitor(CM):

Real-time measurement of computing resources that are observed is restored by Cloud Monitor. CM utilizes ceilometer and Gnocchi Python API, to restore the gathered data from both VMs and compute nodes (e.g. CPU utilization).

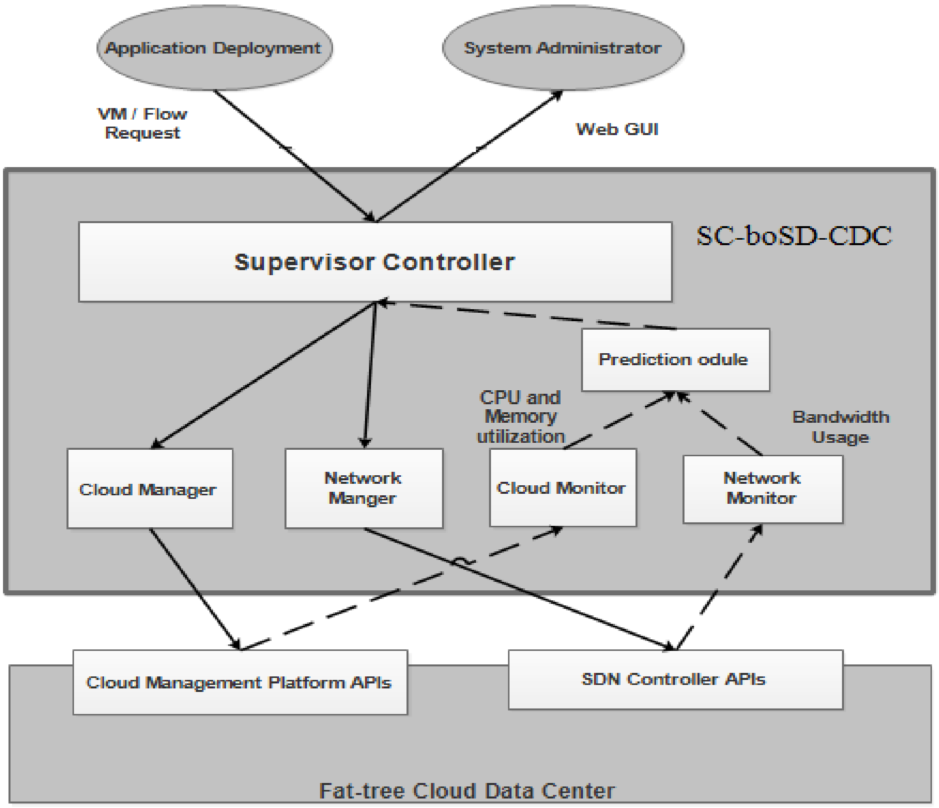

Fig. 4. Data flow schema to components of SC-boSD-CDC.

CONCLUSION

In this paper, we proposed a Supervisor Controller based on the Software-Defined Cloud Data Center (SC-boSD-CDC). The proposed SC-boSD-CDC predicts utilization of compute nodes in VMs and PMs as well as link bandwidth between switches; using prediction data from a multi-objective algorithm to reallocate resources in a CDC under the global view of SDN. In addition, the proposed system reduces overall power consumption in DC. In the future, the proposed framework will be implemented in the real network setting to find the accuracy of energy consumption and maximizing the utilization of the system.

Reference:

1. Foster I, Zhao Y, Raicu I, Lu S, editors. Cloud computing and grid computing 360-degree compared. 2008 grid computing environments workshop; 2008: Ieee.

2. Akella AV, Xiong K, editors. Quality of service (QoS)-guaranteed network resource allocation via software defined networking (SDN). 2014 IEEE 12th International Conference on Dependable, Autonomic and Secure Computing; 2014: IEEE.
3. Fundation ON. Software-defined networking: The new norm for networks. ONF White Paper. 2012;2:2-6.
4. McKeown N, Anderson T, Balakrishnan H, Parulkar G, Peterson L, Rexford J, et al. OpenFlow: enabling innovation in campus networks. ACM SIGCOMM Computer Communication Review. 2008;38(2):69-74.
5. Goth G. Software-defined networking could shake up more than packets. IEEE Internet Computing. 2011;15(4):6-9.
6. de Souza FR, Miers CC, Fiorese A, Koslovski GP, editors. QoS-aware virtual infrastructures allocation on SDN-based clouds. 2017 17th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGRID); 2017: IEEE.
7. Son J. Integrated provisioning of compute and network resources in Software-Defined Cloud Data Centers 2018.
8. Alkayal E. Optimizing resource allocation using multi-objective particle swarm optimization in cloud computing systems: University of Southampton; 2018.
9. Nasim R. Cost-and Performance-Aware Resource Management in Cloud Infrastructures: Karlstads universitet; 2017.
10. Bamini A, Enoch S. Dynamic Scheduling and Resource Allocation in Cloud. International Journal of Control Theory and Applications. 2017;10(3):63-72.
11. Tseng F-H, Wang X, Chou L-D, Chao H-C, Leung VC. Dynamic resource prediction and allocation for cloud data center using the multiobjective genetic algorithm. IEEE Systems Journal. 2017;12(2):1688-99.
12. Vakilinia S. Energy efficient temporal load aware resource allocation in cloud computing datacenters. Journal of Cloud Computing. 2018;7(1):2.
13. Portaluri G, Giordano S, Kliazovich D, Dorronsoro B, editors. A power efficient genetic algorithm for resource allocation in cloud computing data centers. 2014 IEEE 3rd International Conference on Cloud Networking (CloudNet); 2014: IEEE.
14. Hong W, Wang K, Hsu Y-H, editors. Application-aware resource allocation for SDN-based cloud datacenters. 2013 international conference on cloud computing and big data; 2013: IEEE.
15. de Souza FR, Miers CC, Fiorese A, de Assunçao MD, Koslovski GP. QVIA-SDN: Towards QoS-Aware Virtual Infrastructure Allocation on SDN-based Clouds. Journal of Grid Computing. 2019;17(3):447-72.
16. Feng T, Bi J, Wang K. Allocation and scheduling of network resource for multiple control applications in sdn. China Communications. 2015;12(6):85-95.
17. Heller B, Seetharaman S, Mahadevan P, Yiakoumis Y, Sharma P, Banerjee S, et al., editors. Elastictree: Saving energy in data center networks. Nsdi; 2010.
18. Subbiah S, Perumal V, editors. Energy-aware network resource allocation in SDN. 2016 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET); 2016: IEEE.
19. Medved J, Varga R, Tkacik A, Gray K, editors. Opendaylight: Towards a model-driven sdn controller architecture. Proceeding of IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks 2014; 2014: IEEE.
20. Handigol N, Heller B, Jeyakumar V, Lantz B, McKeown N, editors. Reproducible network experiments using container-based emulation. Proceedings of the 8th international conference on Emerging networking experiments and technologies; 2012.