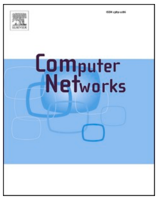
Contents lists available at ScienceDirect

Computer Networks

journal homepage: www.elsevier.com/locate/comnet



Software-defined networks for resource allocation in cloud computing: A survey

Arwa Mohamed [a,e,1,*], Mosab Hamdan [b,1,*], Suleman Khan [c], Ahmed Abdelaziz [d], Sharief F. Babiker [a], Muhammad Imran [f,*], M.N. Marsono [b,*]

[a] *Faculty of Electrical and Electronic Engineering, University of Khartoum, Khartoum, 11115, Sudan*
[b] *School of Electrical Engineering, Faculty of Engineering, Universiti Teknologi Malaysia, Johor, 81310, Malaysia*
[c] *Department of Computer and Information Sciences, Northumbria University, Newcastle Upon Tyne, NE1 8ST, United Kingdom*
[d] *Faculty of Computer Science, Future University, Khartoum, 11115, Sudan*
[e] *Faculty of Engineering, University of Science and Technology, Khartoum, 11115, Sudan*
[f] *College of Computer and Information Sciences, King Saud University, Saudi Arabia*

ARTICLE INFO

Keywords:
Resource allocation
Software-defined networks
Cloud computing
Data Center Network
Edge-Computing
5G

ABSTRACT

Cloud computing has a shared set of resources, including physical servers, networks, storage, and user applications. Resource allocation is a critical issue for cloud computing, especially in Infrastructure-as-a-Service (IaaS). The decision-making process in the cloud computing network is non-trivial as it is handled by switches and routers. Moreover, the network concept drifts resulting from changing user demands are among the problems affecting cloud computing. The cloud data center needs agile and elastic network control functions with control of computing resources to ensure proper virtual machine (VM) operations, traffic performance, and energy conservation. Software-Defined Network (SDN) proffers new opportunities to blueprint resource management to handle cloud services allocation while dynamically updating traffic requirements of running VMs. The inclusion of an SDN for managing the infrastructure in a cloud data center better empowers cloud computing, making it easier to allocate resources. In this survey, we discuss and survey resource allocation in cloud computing based on SDN. It is noted that various related studies did not contain all the required requirements. This study is intended to enhance resource allocation mechanisms that involve both cloud computing and SDN domains. Consequently, we analyze resource allocation mechanisms utilized by various researchers; we categorize and evaluate them based on the measured parameters and the problems presented. This survey also contributes to a better understanding of the core of current research that will allow researchers to obtain further information about the possible cloud computing strategies relevant to IaaS resource allocation.

1. Introduction

Cloud Computing (CC) has recently come out, and it has been viewed as allowing a common collection of configurable computing services to be made accessible and released as specified by the National Institute of Standards and Technology (NIST) [1]. It also allows easy and on-demand network access. Thus, networks, servers, storage, applications, and services resources are pooled in the cloud to serving several tenants. Services providers, e.g., Microsoft Azure, Amazon, and Google Cloud, provide access through the internet to CC resources based on a pay-per-use policy. Nowadays, in a few hours, anyone can pay for cloud services, deploy and sets up servers for an application. The physical infrastructure is leased out to CC clients based on leasing it from an external cloud service provider. Therefore, they only pay for the resources they use.

NIST classifies CC into three operation models and four deployment models. The CC operational models are divided into three specific categories: Infrastructure-as-a-Service (IaaS), Platform-as-a-Service (PaaS), and Software-as-a-Service (SaaS). In IaaS, the underlying cloud infrastructure can not be managed or control by clients. Other than that, clients can set up and run the software, control operating systems via administrative access to VMs and programs, and allocate processing,

* Corresponding author.
E-mail addresses: arwa.eldhai@ust.edu.sd (A. Mohamed), mosab.hamdan@ieee.org (M. Hamdan), dr.m.imran@ieee.org (M. Imran), mnadzir@utm.my (M.N. Marsono).

[1] The first two authors contributed equally to this work.

https://doi.org/10.1016/j.comnet.2021.108151
Received 28 December 2020; Received in revised form 11 April 2021; Accepted 3 May 2021
Available online 14 May 2021

storage, networks, and other essential computing resources. On the other hand, PaaS allows developers to deploy web-based applications without purchasing and setting up physical servers. Meanwhile, SaaS provides host applications and makes them accessible to customers over the internet, e.g., Google, Salesforce, Microsoft, and Zoho. Development models define how customers and cloud providers communicate with each other through private, public, hybrid, and community clouds.

A public cloud is operated by a single corporation and is open to the public or a broad business group. A private cloud is own by a specific organization. Cost-effectiveness, reliability, flexibility, and high scalability are the main benefits of the public cloud. Still, it lacks security compared to a private cloud that guarantees high security but suffers from high cost and limited scalability. A community cloud shares a small number of relationships and builds specific groups with a common duty. An infrastructure that contains several clouds of any kind is called a hybrid cloud, which they communicate between them to allow the transfer of data and applications from one cloud to another through their interfaces.

CC built on data center infrastructure is designed with the ability to apply virtualization technology [2]. Virtualization technologies enable flexible resource allocation to virtual machines (VMs), enabling flexible resource provision on-demand. Moreover, virtualization allows multiple applications to be integrated on fewer physical servers, which promises to be significant cost-savings due to increased energy efficiency and lower system administration costs. Compared to sharing the main computational resources in cloud computing, sharing network resources is more difficult. Conventional routing schemes in Data Center Networks (DCNs), including the Open Shortest Path First (OSPF) [3] and Routing Information Protocol (RIP) [4], are based on static routing; and therefore loses flexibility in identifying flow paths for different network states. Likewise, VM scalability causes some bottlenecks to result from across-VM communication on a single host due to virtual bridges. Furthermore, the conventional routing process limits the efficient use of resource capacity. Hence, it cannot build an improved network knowing that the general cost of fixing and configuration these networks is very high.

Software-Defined Networks (SDN) [5] is an advanced network model that improves the shortcomings of the conventional network infrastructures by separating the control plane and the data plane from switches and routers. SDN technology enables network control through centralized software controllers and makes network management more efficient, fast, and flexible. The SDN central controller has a global network view to deal with the dynamic changes in the network topology. Simultaneously, the SDN controller communicates to the forwarding device through the OpenFlow protocol [6]. The flexibility of SDN-managed networks has opened up new opportunities for the research community to incorporate IaaS and Cloud Data Center (CDC) capabilities into the SDN [7]. SDN also provides resource management systems that allow cloud services provision while attaining dynamic traffic requirements when running VMs. Furthermore, SDN and Network Function Virtualization (NFV) [8] enable much greater network flexibility by splitting network architectures into virtual slices, a process is known as network slicing, which aids fifth generation (5G) wireless technology innovation [9].

CDC systems require agile and flexible network control functions with computing resource control to ensure characteristics of traffic performance for the VM operations are adequate and attainable by the SDN [10]. Also, SDN virtualizes on-demand network resources to utilize resources and satisfy user application restrictions efficiently. Resource allocation has become one of the obstacles of cloud computing arising from users sharing computing and network resources, and the network often adopts the best-effort transmission mechanism [11] and Shortest-Path-First (SPF) routing mechanism [12] for data transmission, leading to a link load imbalance and high possibility of link congestion. Indeed, the network topology and routing mechanism have a major effect on the Service Level Agreement (SLA), the Quality-of-Service (QoS), and the network latency [13] that negatively affect the energy usage of Cloud Computing Data Centers Networks (CCDCNs). In turn, the ineffective allocation of computing resources leads to an overprovisioning or an underprovisioning, negatively impacting the SLA. Consequently, it reduces the profit for the cloud provider and increases the user's cost [14]. Fig. 1 depicts the resource allocation process.

In CC, unpredictable and changing requests of resources among end-users depending on their application usage style are the key challenge of CC. Moreover, resource allocation aims to optimize applications, i.e., QoS, improved resource utilization, and power efficiency, no matter what type of Information and Communication Technologies (ICT) resources are allocated to end-users. Consequently, integrating SDNs into a cloud computing environment solves most of the previously present problems in resource allocation in cloud computing networks. Hence, SDN allows implementing policies, configuring, and managing network resources in a short time per periods one control protocol to perform a series of operations, including routing, traffic engineering, load balancing [15,16], and access control.

1.1. Scope and contribution

Recently, many methods, techniques, and algorithms have been implemented, emphasizing cloud computing based on SDN in resource management, resource scheduling, resource allocation, energy conservation, load balancing, and QoS. This article aims to provide a thorough overview and survey of current resource allocation techniques, frameworks, and models for cloud computing based on SDN. Our contributions can be summarized as follows:

- We put forward the state-of-the-art resource allocation mechanisms for CC based on SDN.
- We are presenting a taxonomy of recent trends in resource allocation mechanization while ensuring their advantages and disadvantages.
- We set down the performance criteria utilized for assessing the current techniques.
- We explain the potential research work that has already been stated, which helps to identify the path for current and future usage.

1.2. Organization

The following paragraphs shall be arranged as follows. The information on cloud computing and SDN and the history of resource allocation in Section 2. Section 3 sets out the motivation for researching and improving cloud computing-based SDN. Section 4 points out the resource allocation in CC based on SDN and the types of resources and parameters used throughout the allocation. Section 5 introduces our resource allocation taxonomy. Section 6 contains a list of works relevant to resource allocation in CC based on SDN. Open study issues and recommendations for future research in Section 7. The article is concluded in part 8.

2. Related Work

This section summarizes the relevant literature on the definition of resource allocation in the SDN, cloud computing, 5 G mobile communication, and edge cloud computing. Indeed, we have not found a thorough survey covering the two fields of CC and SDN, especially in terms of resource allocation. So, we are presenting the state-of-the-art in resource allocations in different fields separately. All related works with the term "Resource Allocation" in the name or acronym reported from January 2014 to March 2021 were selected for the first time from scientific journals, namely IEEE, Elsevier, Springer, and other international journals. The field of these survey papers in comparison to our paper exemplifies in Table 1.

A systematic analysis of IaaS-related resource allocation in cloud computing is presented in a comprehensive research paper [17] on

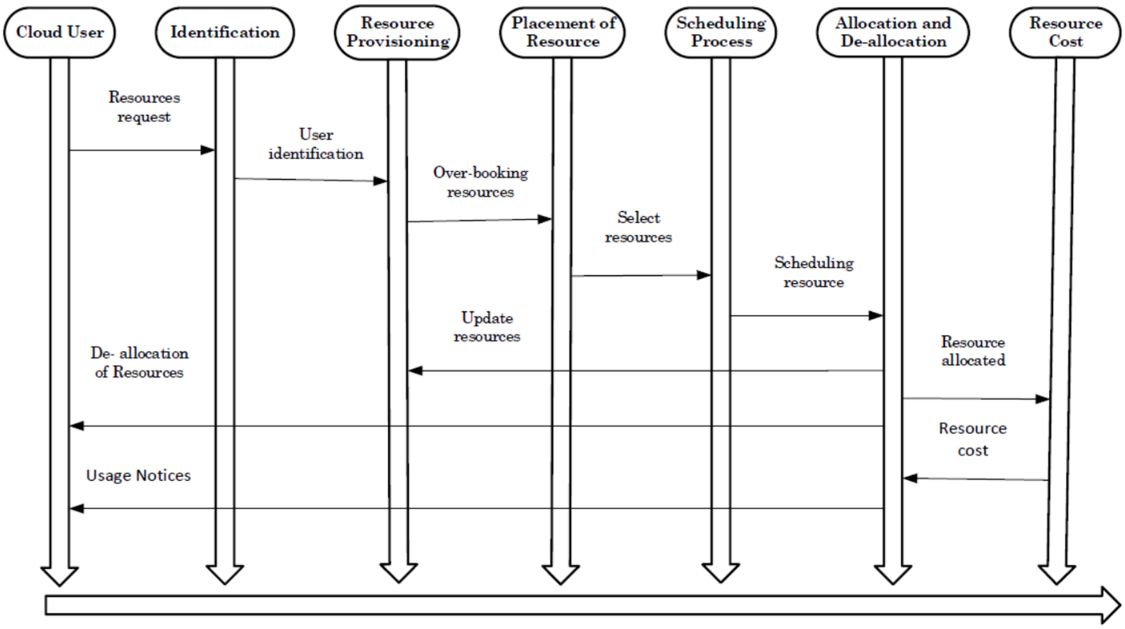


Fig. 1. Resource allocation process.

Table 1
Related works on resource allocation process.

Survey Paper	Year	Taxonomy	Issues and Challenges	Cloud computing	SDN	Edge cloud computing	5G	Future work
This work	2021	✓	✓	✓	✓	✓	✓	✓
Xu [34]	2021	✓	✓	×	×	×	✓	✓
Hamaali [33]	2021	×	✓	✓	✓	✓	✓	×
Saidi [26]	2020	×	✓	✓	×	×	×	×
Ejaz [32]	2020	✓	✓	✓	×	×	✓	✓
Le Duc [35]	2019	✓	✓	×	×	✓	×	✓
Su [36]	2019	✓	✓	×	✓	×	✓	✓
Lakhwani [25]	2019	✓	✓	✓	×	×	×	×
Banumathi [18]	2018	✓	✓	✓	×	×	×	×
Wang [28]	2018	✓	✓	✓	×	×	×	✓
Madni [17]	2017	✓	✓	✓	×	×	×	✓
Zehra [37]	2017	×	✓	×	✓	×	×	✓
Hamdy [21]	2017	×	✓	✓	×	×	×	×
Yousafzai [24]	2017	✓	✓	✓	×	×	×	✓
Akhter [29]	2016	✓	✓	✓	×	×	×	✓
Hameed [30]	2016	✓	✓	✓	×	×	×	✓
Alnajdi [27]	2016	✓	✓	✓	×	×	×	✓
Hamdi [31]	2016	×	✓	✓	×	×	×	×
Jennings [23]	2015	✓	✓	✓	×	×	×	✓
Anuradha [20]	2014	×	×	✓	×	×	×	×
Mohamaddiah [22]	2014	×	×	✓	×	×	×	×
Nunes [38]	2014	×	×	✓	✓	×	×	×
Asha [19]	2013	×	×	✓	×	×	×	×

resource allocation methods used in cloud computing resource allocations. The survey classifies these methods into two main types depending on the problems addressed and evaluation parameters into either the strategy-based or the parameter-based. These two basic types are then divided into other subtypes. It is concluded that many previous studies did not consider important parameters for allocating resources in IaaS that require better improvement in this area. Banumathi et al. [18] have reviewed several policies for resource allocation in cloud computing into three types: static, dynamic, and advanced. These resource allocation techniques were discussed in detail in terms of their features and limitation. In contrast, Asha et al. [19] presented a comprehensive review of resource allocation approaches and their effect on the cloud computing system. The survey discussed the four-point definition of fluctuations resulting from network topology, dynamic configuration, unforeseen resources, and concurrent data processing in the CC environment. The significance, advantages, and disadvantages of each system are also considered. However, different issues concerning resource allocation stay unaddressed. Also, Anuradha et al. [20] have elaborate several resource allocation policies that have been reviewed based on the utility functions, priority-based algorithms, multiple-criteria decision analysis, genetic algorithms, and skewness algorithms. The resource allocation techniques referred to above have been explained in depth with their benefits and drawbacks of each technique. They found that genetic algorithms are the most effective for the allocation process.

Similarly, Hamdy et al. [21] presented many resource allocation techniques and discussed several parameters that impact the used resource allocation policy. Muhammadiyah et al. [22] conducted a review in cloud computing concerned with allocating and monitoring resources. They provided a classification according to the cloud definitions, characteristics, and deployment models. Also, introduce the problems and challenges associated with the resource allocation process, monitor them, find appropriate ways to solve them to attain better performance, avoid the SLA violation, enhance the resource performance and save the power consumption. Reference [23] organized the former research into a cloud management platform, which summarized in five basic points as follows: a) For cloud-hosted applications provide predictable performance; b) Realize global view manipulating for cloud systems; c) Scalability for resource management systems; d) Learn about cloud pricing and its economic behavior, and e) provide for mobile cloud solutions and how to develop them. Also, they have illustrated the pros and cons as well as a set of essential research challenges. Yousafzai et al. [24] demonstrated the strengths and weaknesses of the state-of-the-art cloud resource allocation strategies. They also proposed a stylistic taxonomy focused on optimizing the allocation of resources by categorizing existing literature. Likewise [25] and [26] presented various resource allocation techniques and challenges for efficient resource allocation and efforts to improve QoS.

Dynamic resource allocation was addressed by [27], who ranked them according to their strategies in use-based RA, market-based dynamic RA, and dynamic SLA-based RA. They also clarified their strengths and limitations. Additionally, research directions and findings from the literature review are included. Further, areas that require further research were also identified. Reference [28] discussed the framework and directions related to scheduling strategies and resource allocation algorithms in the cluster frameworks in the data center networks. These algorithms can be classified according to scheduling granularity, controller management, and prior-knowledge requirement. The article also analyzed useful characteristics such as fault tolerance and scalability to shed light on distributed system design principles. In short, all the research work mentioned does not provide a comprehensive survey of resource allocation for CC based on SDN but rather represents the allocation of resources in CC only. In contrast to this work, we present a comprehensive survey of resource allocation for CC based on SDN and an indication of the algorithms and parameters that can assist in this allocation and present challenges and future insights.

Recently, resource allocation gained more relevance to address the issue of power consumption in cloud computing.

References [29] and [30] identified the main points to reduce energy consumption in data centers, maintain the ecosystem, and lower operational costs. These works also illustrated the difficulties and existing methods for solving them. Besides reviewing countless resource allocation approaches in the literature, major open challenges and future research directions were discussed. Reference [29] also discussed efficient VM customization and availability algorithms that must be developed to meet the increasing use of cloud computing. Reference [30] also identified RA problems and appropriate techniques based on the hardware and software available for this purpose. In addition, the RA problems were classified based on different dimensionality as follows: resource adaptation policy, objective function, allocation method, assignment process, and interoperability. It is noticeable that these researches focused only on energy consumption in CDC, unlike our work, as it provides a comprehensive overview of the concept of resource allocation in CC based on SDN.

The allocation of network resources, on the other hand, is one of the most important factors influencing cloud data center activities. Hamdi et al. [31] described network-aware VM placement. This research looked at network allocation in two ways: single and distributed data center cloud, and it divided these two areas into subgroups such as initial placement, migration methods, and both of them. The authors sanction all future challenges and opportunities that take network placement traffic into account when allocating VMs to PMs.

The development of information and closely integrated distributed systems, including the fifth-generation (5G) mobile communications technology, Internet of Things (IoT) applications, edge cloud computing, and big data systems, is accelerating the demand for new frameworks for resource allocation to implementation and testing of large-scale software systems in the cloud computing environment. There are many comprehensive studies on how to allocate resources in these areas, e.g., Ejaz et al. [32] introduced a generalized structure for allocation of resources analysis in CRAN networks. The aim is to provide a detailed survey of resource allocation strategies that could give a CRAN a broader picture of aims, challenges, problem types, and possible solutions. They are addressing several new CRAN usage scenarios, as well as application-specific goals. In the background of 5G and beyond infrastructure, they also reviewed problems and critical problems in a CRAN. Moreover, the distributed resources in terms of allocation are described while keeping in mind the concerns of providing millions of consumers without lack of success or failure. As a result, a detailed analysis of resource allocation for distributed systems using descriptive algorithms was performed. Regardless, future studies need to be conducted to perform more advanced algorithms in each area [33].

From the standpoint of using resource allocation to achieve spectrum processing and reduce cross-tier interference, reference [34] presented a comprehensive overview that was dealt with resource allocation algorithms in modern HetNets for 5G. The focus of this research is on various HetNets network scenarios. Consequently, they introduced the categorization of current RA algorithms and discussed some urgent questions and potential research directions. This paper addressed two possible 6G communication architectures for addressing next-generation HetNets RA concerns: learning-based RA architecture and control-based RA architecture. Le Duc et al. [35] express a type of machine learning implemented to detect intrusions of reliable resource provisioning in joint edge-cloud environments, as well as a survey of techniques, frameworks, and methods which can be used to increase the reliability of distributed applications in a wide range and heterogeneous distributed systems. The survey is organized around a technical breakdown of the credible resource allocation issue into three main categories: workload characterization and prediction, component placement and system consolidation, and application elasticity and remediation. To conclude the article, a summary of major obstacles and a highlight of recommendations for future research are presented. Each of these previous studies offered a thorough examination of the various fields, depending on the feature of resource allocation. In contrast to our work in this paper, these aforementioned publications do not include the allocation process for both the CC and SDN domains.

In the sense of using SDN for allocating network resources, a survey of the latest developments in network resource allocation is presented by Zehra et al. [37] They introduced techniques for research allocation in SDN and discussed all advantages and disadvantages of each technique. Likewise, network slicing is mainly influenced by SDN and Network Function virtualization (NFV). Su et al. [36] presented a survey on resource allocation algorithms in 5 G network slicing in terms of concepts and mathematical models. According to the research goals, the mathematical models for resource allocation can be divided into four groups: general, economic, game and prediction, and robustness and failure recovery. Each models inspiration, purpose, and key concept are explained and evaluated using the most recent examples. These previous studies focused only on the concept of allocation in SDN compared to our work in this paper, including allocation in two fields: SDN and cloud.

One of the scientific papers that dealt with the topic of resource allocation in CC based on SDN is by Nunes et al. [38] to investigate new dynamic resource allocation strategies and their key characteristics and to summarize the key developments in the fields of VMs and CC using SDN networks and virtual networks. On the contrary, we present our latest CC based on the SDN resource allocation mechanism, comparing previous research in this field. In addition to providing a classification of recent developments in resource allocation mechanization, we also assess their benefits and drawbacks. We also take a look at the performance metrics and algorithms that are used to assess existing technologies. This work also explains potential research work that has already been discussed, assisting in establishing a route for current and future use.

3. CC based on SDN

Cloud computing [39] is a successful and evolving model for delivering ICT resources as services for both governments, educational and industrial fields over the internet. Cloud service providers provide three forms of application services: SaaS, PaaS, and IaaS. Cloud users can conveniently rent these services as they pay by the size of usage.

CC has dozens of servers that are connected to thousands of switches. Virtualization technology has been included in CC to facilitate scalability and flexibility. In other words, virtualization technology is used for virtual computing nodes, storage, and network resources, and it provides these resources as services to cloud users. Consequently, cloud consumers can lease these virtual computing resources, i.e., VMs or storage, from cloud providers. The success of computing resource virtualization resulted from using well-defined abstraction mechanisms that simplify virtualization operations [40,41]. However, implementing the concept of a virtual network in cloud computing is a long way off. Especially, existing network virtualization technologies, i.e., a virtual local area network (VLAN) [42] and virtual private network (VPN) [43] do not provide adequate solutions.

There are multiple VMs in the same server and consumers can rent any number of these VMs unless the user request exceeds the server capacity. Communication between VMs and the internet is through a variety of routers and switches. The usage and requirements of the network are growing at a very rapid pace and, to meet current demands, there is a need to automatically increase the size of the infrastructure. From this point, the application of conventional networks to cloud computing networks is very difficult and complicated, as it is considered time-consuming and costly, especially in the state of VM migration and network configuration.

Operators can adopt SDN in traditional cloud data centers to

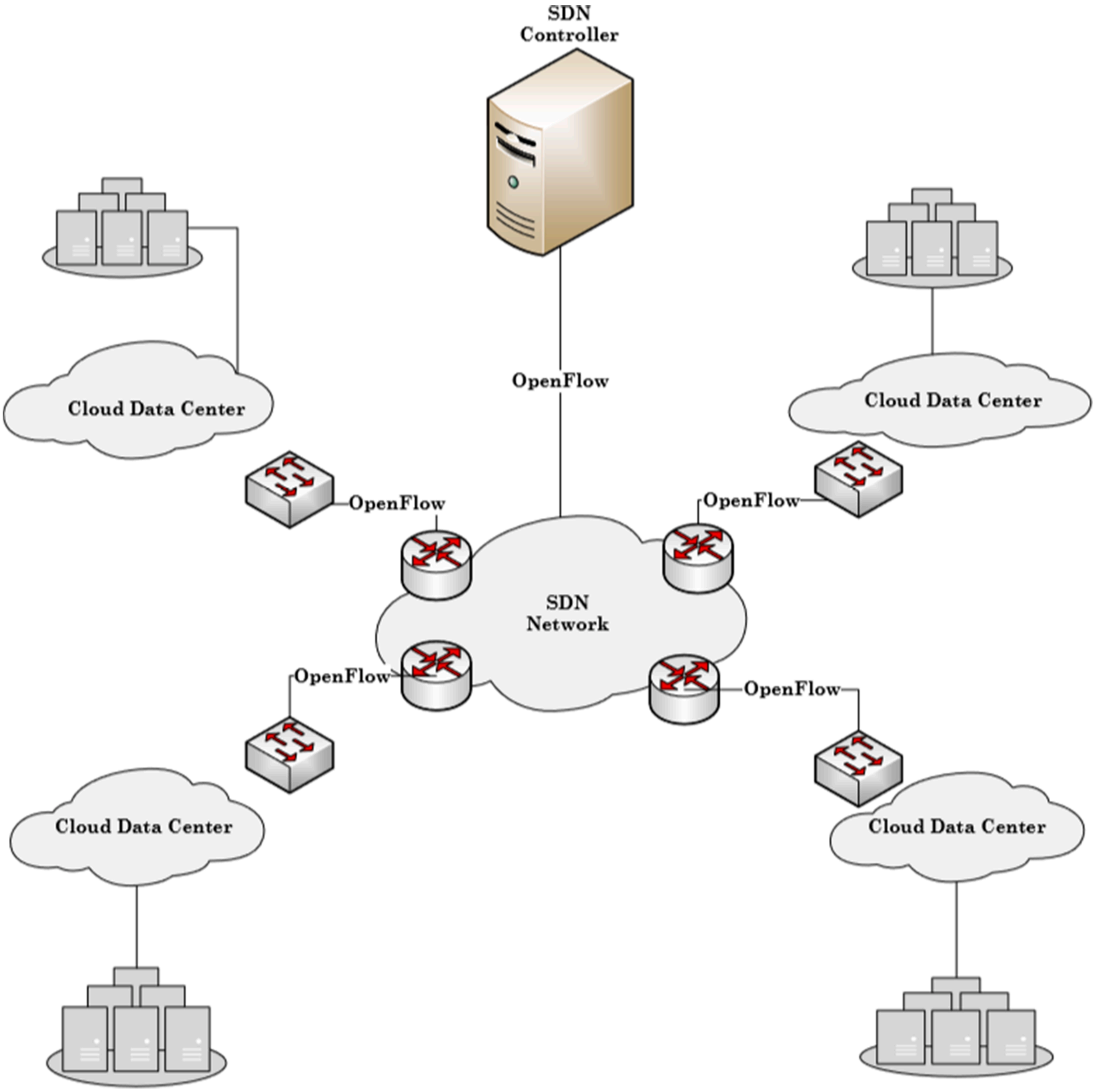


Fig. 2. Cloud computing based on SDN architecture.

alleviate challenges facing their data center networks [44]. An SDN allows network configuration control routes to be separated from network devices. This capability provides flexibility to the network plane control by easily adapting to changes in the network through a program called a controller. Controllers communicate with the forwarding plane through OpenFlow, allowing them to make changes occurring to the network, hence achieving real-time response to traffic requests. In short, SDN provides efficient, flexible, agile, and scalable solutions for CDCNs. Besides, CDC significantly improved efficiency when SDN adapted in their environment in both the performance of the network [45,46], security [47], network availability [48], and improved energy efficiency [49,50]. Efficiency can be further improved by providing QoS for applications by slicing the network and allocating appropriate dynamic bandwidth [51,52]. Cloud computing based on SDN architecture is illustrated in Fig. 2. Worthwhile, Google's cloud provider officially implemented the SDN model in its data center to promote manageability and improve scalability [53].

4. Resource allocation in cc based on SDN

The difference between the three concepts of resource provision, resource allocation, and resource scheduling is explained in [54] as follows: the process of resource provisioning is when a service provider allocates resources to a consumer, while when dividing out resources between competing sets of programs or clients economically in this case called resource allocation. On the other hand, resource scheduling is a timeframe for allocating resources in which the resources are gathered and made accessible at specific times, and the computational procedures are coordinated at that time. In other words, the allocation of cloud resources is a systematic process, including resource discovery, collection, provisioning, application planning, and resource management.

The allocation process can be detailed as a way of identifying, choosing, supplying, scheduling, managing, and assigning available computational, storage, networking, and energy resources to a set of applications over the internet, to achieve the common objectives of each cloud application, user, and provider. These goals differ from each other and are subject to criteria such as application requirements, SLA, over-utilization, and under-utilization of resources. SLA is considered a contract between the users and its providers. This contract specifies how to provide resources to ensure the achievement of the goals of each party. The allocation of resources to a cloud computing data center is a key problem for optimizing computational resources, network resources, and energy usage.

Ineffective allocation affects the performance of the entire cloud environment. In a CDC, cloud providers play an important role in the resource allocation process as they provide resources in the form of IaaS, PaaS, or SaaS per the SLA agreed between all users/consumers of the cloud and to achieve other management goals. Noteworthy, realizing a high use of resources to maximize revenue is the desire of the cloud provider, while consumers want to decrease expenditure while achieving their performance requirements. However, due to the lack of information sharing between the parties, optimally distributing resources is not trivial. Furthermore, the optimally physical location, dynamic fluctuation of the environment, and unpredictability of resources in the nodes that cannot be attained with traditional resource allocation form harder challenges for both actors.

Cloud resources (virtual or physical) are located in data centers and are allowed to share between a group of users and must be dynamically allocated and modified on-demand. For example, a consumer can request a networked resource, such as bandwidth and delay, or a computing resource such as a central processing unit and storage. Allocating resources in an effective way, while taking into consideration effectively demands and dealing with them in a flexible dealing with them unlimited, thus avoiding the situation of excessive and scarcity of resources in the CDC.

However, sharing network resources in the CDC remains an issue [55,56]. So, we need a comprehensive model for allocating computing and network resources. For that reason, it is important to extend the network infrastructure in an automated way to meet the current and growing demands daily due to big daily data from mobile Internet, multimedia-rich applications [57–59], and the IoT [60–62] that are continuously being collected and processed. The concept of conventional networks that allow switches and routers to make decisions is an ineffective way of dealing with the massive number of these data, specifically in VM generation, migration, and network configuration. To solve these problems, we need efficient, flexible, agile, and scalable networks [63]. Adaptation of SDN in the cloud data center becomes indispensable, offering a brand-new direction for conventional network architecture to construct infrastructure.

SDN works by separating the data plane from the control plane. This separation allows the network centers to be programmed and reconfigured according to the changes on the network. Besides, SDN provides a solution for allocating VMs, network bandwidth. Moreover, the SDN enables various levels of abstraction and automation clarity that direct the individual creation and development of software-based and network-based control systems. Furthermore, the centralized controller of the SDN can collect consumer requests and applies resource allocation algorithms, then forward allocation commands through the network. This can facilitate the resource allocation process in the CDCN. Integration of SDN in the CDC leads to facilitate the process of VM built and deployed, taking into account appropriate resource allocation to avoid uneven server load and all aspects related to QoS and power consumptions. However, this adaptation adds several limitations to the CDC that must be carefully considered [64]

- ***Reliability***: The use of a centralized SDN controller affects reliability.
- ***Scalability***: The SDN controller is becoming a big hindrance as the number of switches and final hosts in the network increases.
- ***Visibility***: SDN only enables the corridor source and the UDP traffic end device to be visible and covers up the user's identity.

4.1. Type of resources in CDC

Fig. 3 illustrates the kinds of CDC services that are designated by the provider to be shared between users. These services, which are provided by a cloud service provider, can be divided into two main types: hardware and software. The software represents applications that are provided to the cloud consumer in the form of PaaS or SaaS [65]. Hardware refers to any of the computing and network resources owned by the cloud provider, which is referred to as IaaS. The following is a summary of these resources. Note that only IaaS is covered in this paper.

A. Computing resources

Compute resources is a combination of Physical Machines (PMs) that includes one or more processors, memory, network interface card, and input/output that jointly provide computational capabilities to the cloud environment. The concept of virtualization applies to PMs, allowing multiple virtual machines to be hosted on a single machine while providing complete separation between them [66], which can run multiple operating systems and applications in the same PM. The optimum use of these resources achieves QoS and avoids violating the SLA, in addition to making a profit for the cloud service providers.

B. Storage resources

Storage resources are crucial in CC technology, as CDCs hold large quantities of computing resources, typically up to millions, and store petabytes and even exabytes of data. Attention must be paid to ensuring that CDC availability, reliability, fault-tolerance, scalability, and energy of distributed data storage can be accomplished in the CC environment [67]. Amazon, Azure, and Google cloud computing providers present a range of storage facilities in cloud data centers, such as virtual disks,

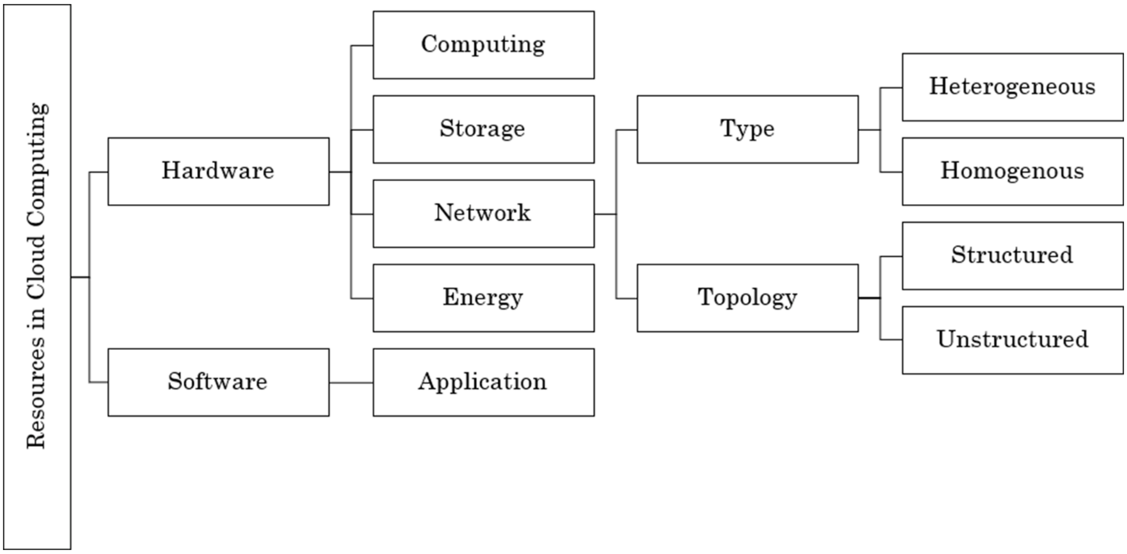


Fig. 3. Computational and network resources in cloud data center.

database services, and object stores. Each of these resources offers different service levels in terms of guarantees of data consistency and reliability.

The most important problems facing these resources are the elasticity represented in the scalability process and decreases according to the dynamic demand of users. It is exceedingly difficult to have high data consistency, atomicity, consistency, isolation, and durability (ACID) properties in legacy networks. As a result, a community has been created called "NoSQL" data storage techniques, enhanced for various organizational and functional circumstances [23].

C. Networking resources

PMs are connected in data centers with many switches and routers with high bandwidth, usually based on Gigabit Ethernet or InfiniBand technologies. These devices are organized into groups placed on racks containing a group of osts to allocate the rest of the allocation technology. SDN led to the emergence of virtual networks, which contain virtual network appliances and virtual links [64]. In a previous point, it is defined as any network device that is not present in a purely physical form but acts as a similar physical equivalent. In contrast, the last point is defined as every traffic flow that uses a custom link referred to as a virtual link. With this in mind, communications overheads imposed by data center networking systems and network protocols restrict the total efficiency of the network. The network can be classified according to type and topology. The network type is determined by the hardware used within the network, whether they are homogeneous or heterogeneous [68]. The topology of the network is divided into structured and unstructured topology [69].

D. Energy resources

Several sources within cloud data centers are considered energy-consuming, such as servers, network equipment, power supply equipment, conditioning infrastructure, and support services (lighting, etc.) [30]. Data centers are operated by one or more energy providers shifting through localized energy generation/storage from alternative electricity, like wind and solar. Most of the current research tends to reduce the use of energy because it greatly contributes to increasing public expenditures for cloud providers and environmental pollution.

According to Ilager et al. [70], the cumulative energy used by data centers in the United States has reached 2%, equal to 70 billion kilowatt-hours of total energy output. In addition, data centers also significantly increase carbon emissions as a result of greenhouse gas emissions. To be precise, it has been found to generate 43 million tons of carbon dioxide annually and continue to increase at an annual rate of 11%. Data Center energy use could be decreased from the worst-case forecast of 8000 Tera Watt per hour to 1200 Tera Watt per hour by 2030 if required. Improving energy efficiency in a CDC is important for cost-effective and sustainable CC.

4.2. Parameters for resource allocating

It is very important to define and evaluate cloud services, so we resort to agreements between provider and user, in addition to QoS parameters that measure the quality of services provided. Whereas the process of allocating resources always seeks to improve these parameters. Fig. 4 shows the proportions of the parameters used in the resource allocation in CC based on SDN. A summary of these parameters as follows:

Response time: Minimum time to respond to a service request to perform the task [71].
Reliability: The ability to successfully complete the runtime [72].
Performance: The number of tasks performed on the request of cloud users [73].
Execution Time: It is also defined as completion time, which is the time taking to satisfy the demands of cloud users [74].
Workload: The amount of processing to be done out for a particular amount of time. It's the ability to process cloud computing jobs [75].
Utilization: The overall amount of resources currently used in data centers. Cloud computing involves maximizing the use of resources to optimize the revenue and income of cloud providers to the satisfaction of cloud users [76].
Throughput: In cloud computing, the total number of tasks fully performed is within a given period [77].
SLA: This is an agreement that describes the QoS offered by cloud providers to cloud users. The Cloud provider is committed to delivering its best service can serve the need of a cloud Customer and avoid violating the SLA [78].
Power: The VM placement & migration strategies used in the cloud data center must reduce their consumption [79].
Fault-tolerant: The system should continue to provide service in spite of the failure of resources.
Cost: The amount to be billed for the use of CC facilities. This is an expense to cloud customers and a benefit and income to cloud providers [80].

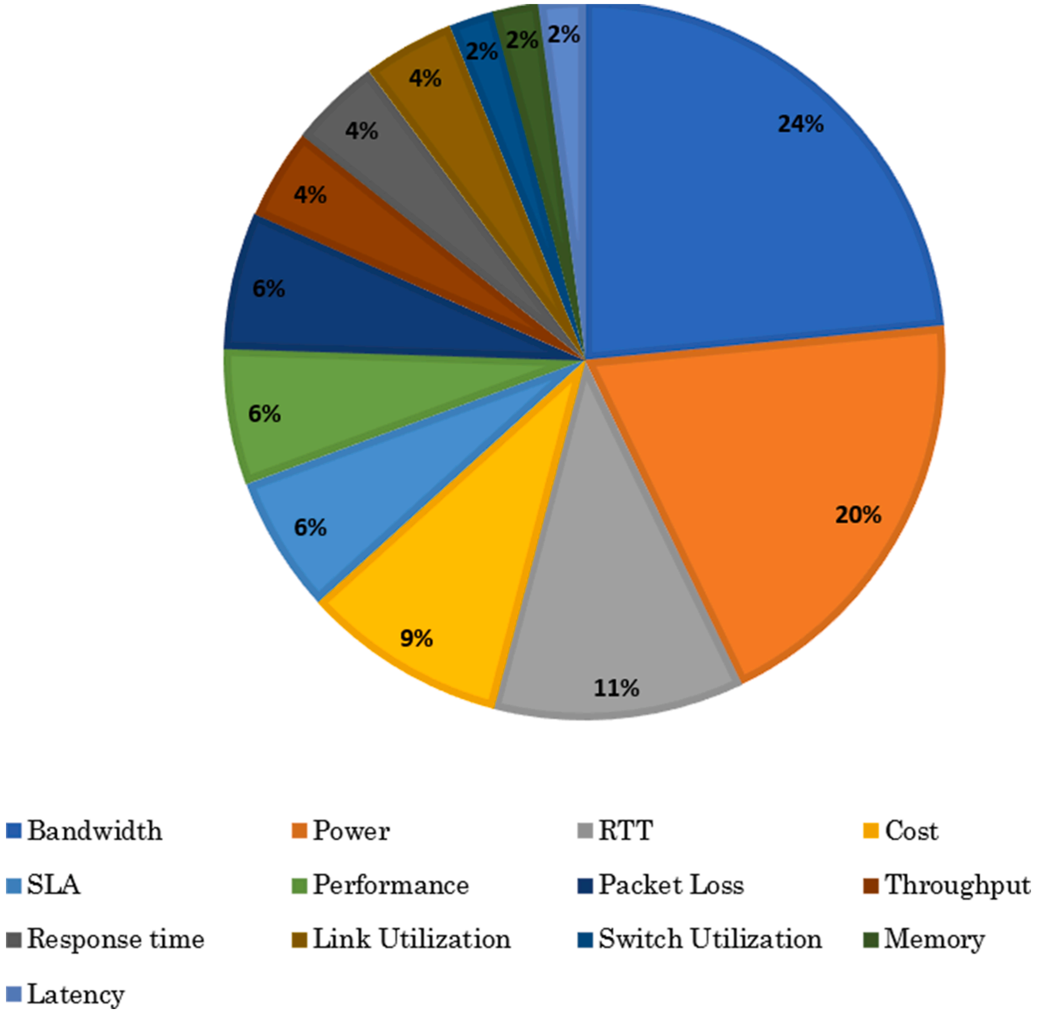


Fig. 4. Parameters used in Allocation.

Bandwidth/speed: Maximum data transmission rate of the network links [81].

Availability: In cloud computing, it represents a collection of services that allow accessibility, maintenance, reliability, durability, and serviceability of the resources that depend on a request of cloud consumers to perform the specified or necessary activity [72].

5. Taxonomy of resource allocation techniques in CC based on SDN

In various research areas, several methods for using CC based on SDN have been proposed. We conducted a thorough investigation into cloud computing using SDN and proposed a classification (shown in Fig. 5) to capture the various aspects of SDN use. Taxonomy is discussed concerning the optimization goal, scope method, evaluation techniques, and optimization methods.

5.1. Optimization goal

The network optimization goal, which includes resource utilization and QoS-aware traffic management in existing data networks, is usually oriented towards either optimizing throughput in congested networks while maintaining adequate transmission quality or balancing traffic to retain a potentially large free capacity for carrying additional (new) traffic [82–85]. In the former, efficient *resource utilization* [15,86,87] is intended to achieve power savings in the CC environment as well as for environmental safety and reduction in data center operational expenditures generate monetary benefits to providers along with environmental harmony. In contrast, *QoS* in cloud computing considers the attainment of user-defined performance metrics in the last point. It can lead to violations of agreed-upon service performance levels when combined with performance metrics. The SLA contracts between a cloud user and a cloud provider specify QoS metrics and specifics in this case.

5.2. Scope method

Here, we propose a taxonomy based on the scope of methodology. As we discuss CC based on SDN, all methods in our taxonomy are fundamentally targeting the following: Joint Optimization-Aware Resource Allocation, Network- aware Resource Allocation, Application-Aware Resource Allocation, and Edge-Aware Resource Allocation.

A. Joint Optimization-Aware Resource Allocation: This method aims to optimize both computing and network resources simultaneously. Recently, one of the biggest problems facing researchers is energy optimization [88] and network performance [89] in a CDC. It is worth noting that power optimization that deals with servers and DCN separately causes limitations on the network because consolidation of a server without taking into account the DCN could lead to traffic congestion and lower network performance. From this perspective, the joint optimization of all resources in CDCs based on SDN is used to overcome the problems mentioned above.

B. Network-aware Resource Allocation: The network-only approach's

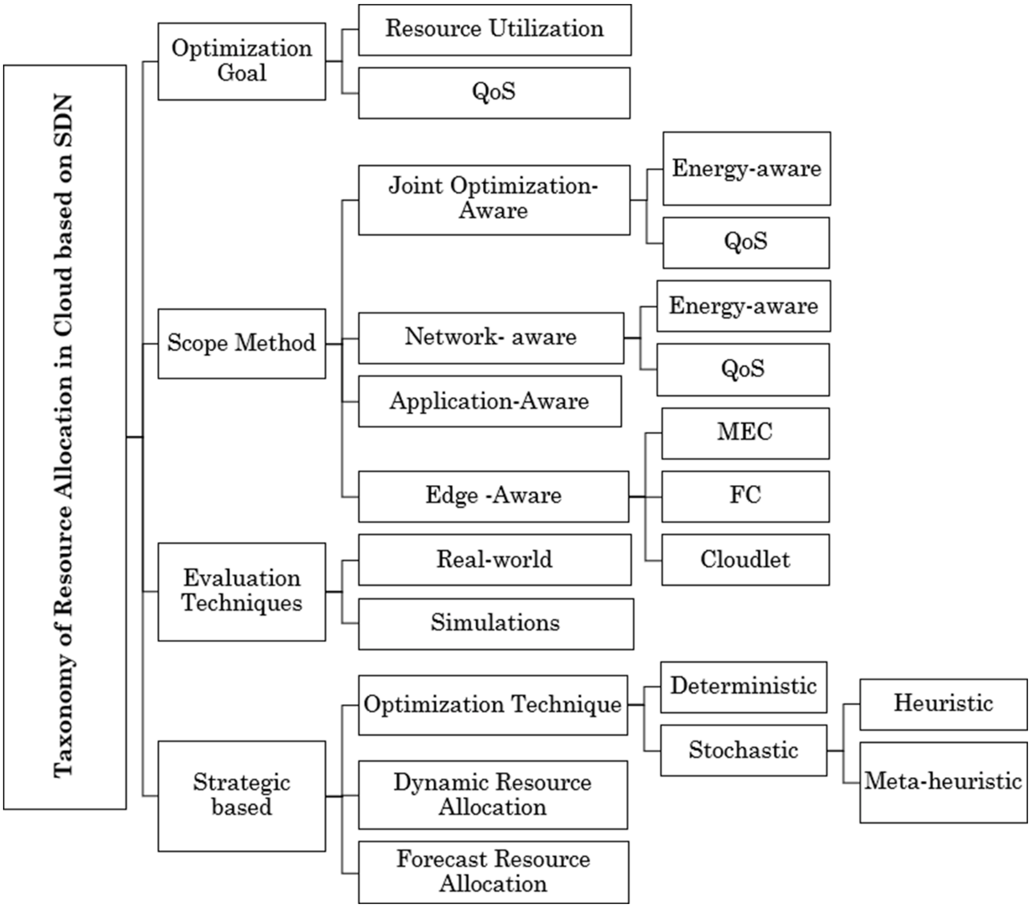


Fig. 5. Taxonomy of resource allocation in cloud based on SDN.

scope is limited to cloud computing's networking capabilities, with no consideration for servers or VMs. These techniques gather network data and use SDN to modify redirect policies in the clouds to address the resource allocation problem.

*C. Application-Aware Resource Allocation: It d*escribes how to delegate applications and resources to CC. Elastic and cost-effective services are provided for most new internet applications by using the infrastructure of CC [90]. Hence, handling various requests of the cloud user application is considered a problem in CC, especially are effect totally in QoS in cloud computing. As a result, the application-aware resource allocation in cloud-based SDN has been addressed by researchers who rely on load detection in the application and assign VMs or links and various SLAs to the applications.

D. Edge-Aware Resource Allocation: The number of IoT devices is projected to reach approximately 125 billion through 2030, while the number of Machine-to-Machine (M2M) communications that represent a large proportion of IoT applications is estimated to cross nearly 45 percent of the national network activity in 2022 [91]. So, the CDC can't handle problems that arise regarding geographical distance, which represents the cloud data center away from the end consumers. This affects network performance and includes access delays, traffic loads, and work costs [92,93], and energy consumption. It was reported that the proximity of network nodes to users effectively contributes to the resolution of the problems mentioned above, hence the term Edge Cloud computing [94,95]. Edge Cloud computing [96] also called, Intel's Intelligent Edge, or Microsoft Cloudnet is introduced to address these requirements of computing, storage, and networking services through highly virtualized platforms situated at the edge of a network. In terms of devices serving as intermediate edge nodes, communication protocols, networks used by the Edge layer, and services offered by the Edge layer, the Edge Layer is implemented in various ways between end devices and the cloud. Mobile Edge Computing (MEC), Fog Computing (FC), and Cloudlet are three types of edge and layer implementation [97]. From this perspective, some researchers focus on this new concept and how to allocate resources in this technology with the help of SDN [98,58].

5.3. Evaluation techniques

The assessment procedure is carried out by the researchers in the form of real-world implementation, simulation, or both to evaluate performance.

A. The implementation: Is usually done in CC based on SDN using real or virtual servers and resources. While this outcome is even more precise, assessing a huge scale on real-world systems is expensive because the price will be prohibitively high, and the management of CDC components can be extremely difficult.

B. Simulations: As opposed to real-world implementation, can save money and allow for more flexibility in management. However, since simulation results are statistically calculated from certain settings, the simulation result can be inaccurate if the configuration varies too much from the validated environment. Simulators, such as CloudSimSDN [99] and a hybrid platform of Mininet [100] and POX [101] suggested by Teixeira et al. [102], are used at the same time to analyze CC based on SDN. Therefore, Openstack [103] and OpenStackEmu [104] are used for real implementation evaluation to attain computational and network resources in both cloud and SDN simultaneously.

5.4. *Strategic-based resource allocation*

On the basis of technique behavior and environment, strategic-based resource allocation is further divided into three groups: optimization technique resource allocation, dynamic resource allocation, and forecast resource allocation. The following are the specifics of the categorization mentioned above:

A. Optimization Technique Resource Allocation: Concerning problems arising from various objectives, resource allocation is a major problem that needs to be improved. As a result, the optimization strategy method that allocates resources using artificial using intelligent algorithms that act and behave like humans is a must. modeling and evaluating solutions based on an objective function, then applying research methods to find the best solutions, are the responsibilities of optimization algorithms. Depending on the number of goals included in the evaluation methodologies, the goal function can be computed using a single goal or multiple objectives optimization. Many intelligent methods have been developed, including deterministic and stochastic algorithms, to improve resource allocation in CC based on SDN, depending on the details of the algorithm and the approach for solving problems. Deterministic algorithms use a predictable path and variables, while stochastic algorithms use unpredictability in the direction and variables. There are two kinds of stochastic algorithms: heuristic and meta-heuristic. Heuristic methods find a suitable optimum solution with low computational cost, but they aren't guaranteed to do so [105]. Meta-heuristic algorithms, which combine randomization and local search, outperform simple heuristics in most cases [106].

B. Dynamic Resource Allocation: The fluctuating demands of cloud users are one of the issues that exist in CC based on SDN, so dynamic resource allocation in cloud computing must satisfy the allocation based on the changes that occur in the system on a regular basis. Dynamic resource allocation techniques are used to handle and meet these unpredictably high demands based on user needs in various scenarios and workloads. It also entailed securing QoS in order to prevent SLA violence. This personalization approach is used in a variety of strategies by researchers [107–110].

C. Forecast Resource Allocation: Predicting future user demand, influencing resource requirements, and automatically assigning resources are all considered important in cloud computing resource allocation [111,112]. Forecasted resource allocation is used for such reasons to allocate or reserve resources for the upcoming before they are requested. It is important and necessary for effective resource management in CC based on SDN.

6. Resource Allocation in cloud computing based on SDN: Current research

Following the classification proposed in Section 5, we present the related surveys published in cloud computing based on SDN in this section. Using the taxonomy outlined in the previous section, we categorize the most recent research based on the paper's main contribution. In this section, we use the survey's main goal as the basis for classification. The subgroups employed for more classification of the studies surveyed in this portion are shown in Fig. 6. The literature for each category is described in detail below.

6.1. *Joint optimization-aware resource allocation*

Recently, one of the most pressing issues confronting researchers has been how to develop both computing and network resources at the same time without affecting the other. Fig. 7 illustrates the Joint Optimization-Aware Resource Allocation and the parameters used in each case. The researchers' point of view on Joint Optimization-Aware Resource Allocation is split as follows:

A. Energy-aware resource allocation

Energy efficiency is an important and fundamental research issue in CC. The energy efficiency of CDCs using SDN technology is improved by reducing the energy consumption in VM and network using placement, consolidation, and overbooking techniques [113]. The initial placement of VMs, migration, and the consolidation of VMs into fewer servers contribute to an optimization of the total power usage of the data center

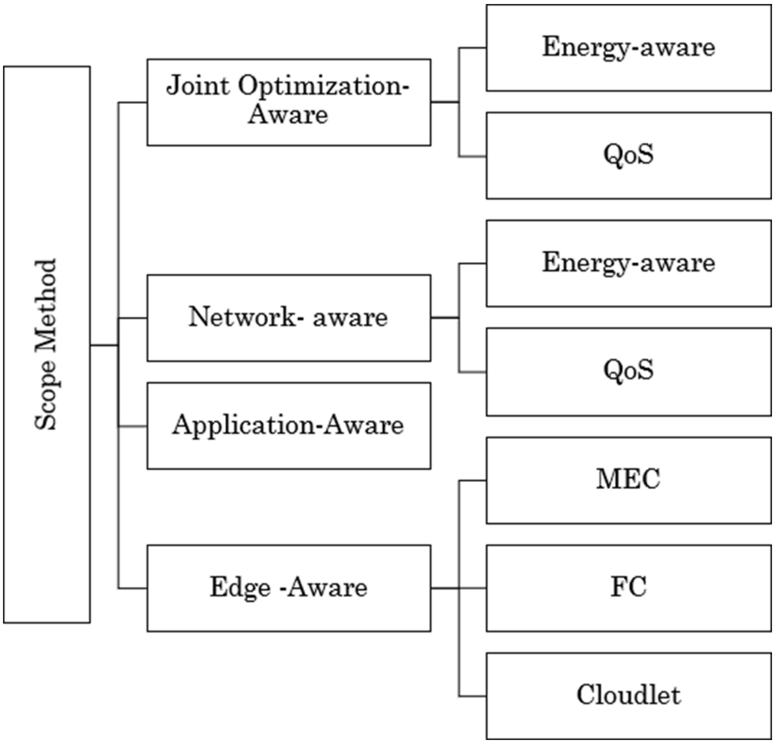


Fig. 6. Studies SURVEYED SCOPE.

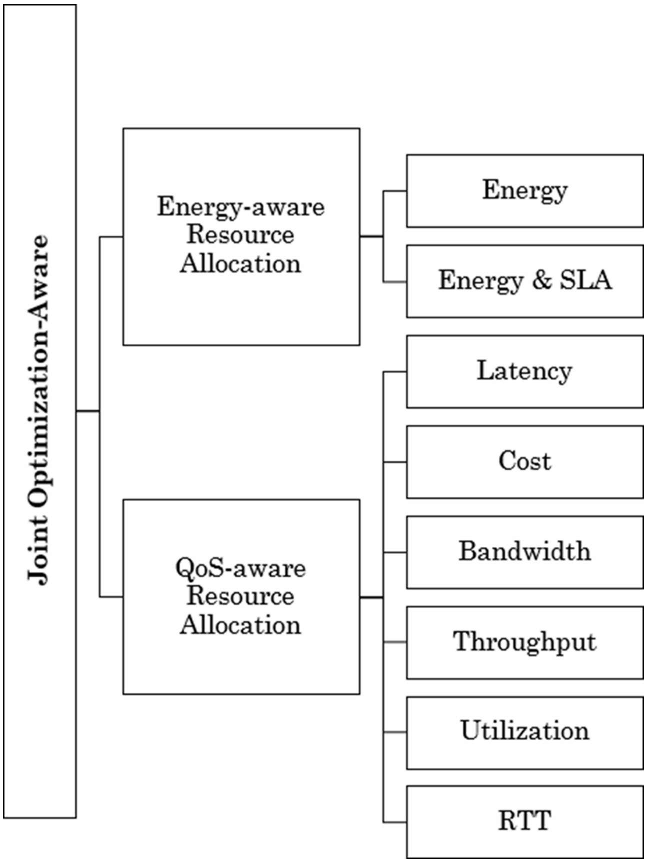


Fig. 7. Joint optimization- aware resource allocation parameters.

as a whole. Adopting SDN into IoT, especially machine-to-machine communication, facilitates smart energy management, especially in resource allocation mechanisms [114].

Additionally, the overbooking techniques, which put more than VMs into a host to optimize resource utilization, often contribute to power saving. Hence, the integrated methods are then used to take into account both the network and the servers simultaneously rather than doing the optimization process separately.

To attain a trade-off between power-saving and fulfill user QoS requirements while eliminating the network congestion, the QRVE mechanism based on distributed ODL controller is propose by Habibi et al. [115]. Therefore, QRVE partitions the fat-tree topology into clusters and applies the VM placement algorithm to fulfill users; SLA and saves power in the DCN. Then, the routing algorithm is applied to the current DCN topology to consider the elephant flow technique to avoid congestion. Even so, the mechanism does not use the migration technique and applying it in the emulator environment. Likewise, Son et al. [116] formulated SLA minimization and power reduction as a multi-commodity problem; therefore, they dealt with each one separately. They applied dynamic overbooking policies to dynamically allocate a host and network resources by monitoring historical real-time workload utilization using correlation analysis. They took advantage of virtualization abilities and SDN for VM placement and traffic consolidation to implement this work in a data centers homogeneous configuration. Lin et al. [117] presented a unified solution containing flow migration and VM migration for the SDN based cloud data centers. Two techniques accomplish this work: traffic-aware flow migration with a dynamic-reroute algorithm (DENTIST) into DENDIST-FM for SDN and energy-and-topology aware VM migration (ETAVMM) to enhance power conception as well as network performance. The unified solution is tested and implemented using Network Simulator (NS2) [118] v2.34 and CloudSim v3.0 [119].

In optimizing resource efficiency and reducing energy usage, the prior concept relies on VM placement, irrespective of the inherent traffic between VMs. However, designing power-saving routing and flow scheduling in the latter process, neglecting resource requests in VMs. Therefore, joint optimization by utilizing VM placement and flow routing is considered. For these purposes and to simplify the joint problem, Jin et al. [120] transformed the VM placement problem into the routing problem. Thus, the authors used the clusters mechanism to attain a fast completion time of large SCALE VMs in DCN. The depth-first best-fit routing algorithm is also utilized to quickly find the paths for both host and network to maximize flow consolidation. The proposed approach is evaluated in both simulation and real implementation using a fat-tree topology-based OpenFlow system. Furthermore, Yu et al. [121] focused on allocating VMs and interconnection between them by implementing a new heuristic algorithm from the Data Field. However, VMs migration and traffic congestion aren't considered in the improvement.

Zheng et al. [122] present PowerNetS to save energy based on important observations by correlating the workloads of various hosts and DCN traffic flows throughout server consolidation. PowerNetS results in lower cross-server traffic and thereby lower power consumption and reduced network delays. It is worth noting that this work utilized homogeneous servers and did not consider VM placement.

SDN assisted Virtual Data Center (VDC) embedding solution called SAVE [123] is a solution that allocates VMs among multi-data centers. It was formulated based on three key technologies: VDC embedding, dynamic traffic engineering, and Locator/Identifier Separation Protocol (LISP) based VM live migration. Hence, SAVE can discover the best VDC components mapping and the optimal routing paths on various data center environments; to lower power consumption to maximize the revenue of cloud providers. SAVE has been tested and implemented with three heuristic algorithms to solve a VDC embedded problem using the Mininet emulator.

On the other hand, Liao et al. [124] focus on designing a hybrid partheno-genetic algorithm to fix the energy consumption optimization in CDC based on SDN. Subsequently, the proposed algorithm addresses VMs migration and integration of CDC by considering flow bandwidth between VMs, therefore reducing active servers and switches in DCN, which reduces the whole cloud data center equipment that consumes power. Various energy-aware resource allocation mechanisms are compared on different scales, and the main differences are listed in Table 2.

B. QoS-aware resource allocation

A few of the challenges that cloud computing applications face are related to QoS management, which is how to allocate computing and network resources that are appropriate for each application. Network virtualization was one of the solutions suggested. Network virtualization is to partition physical network resources such as storage devices, operating systems, or any processing factor in CDCs into smaller segments and rent it to cloud tenants such as cloud-based VM allowed by host virtualization. Several authors have introduced methods that use virtualization technology to dynamically and statically map available resources while optimizing the number of servers used, tracking application requests, and supporting green computing.

Reference [125] formulated a method to allocate adequate resources to high-priority applications in multi-tenant cloud data centers that meet QoS requirements for the application. The authors proposed a combined algorithm contain a priority-aware VM allocation algorithm and bandwidth-allocation algorithm, which are used to allocate computing and networking resources for high-priority applications even in the case of cloud computing is busy. The work was implemented using CloudSimSDN for both synthetic and real (Wikipedia) workloads. Allocating virtual infrastructure in the cloud data center has become a challenge despite implementing SDN in its infrastructure. In contrast, SDN has demonstrated other challenges related to limiting flow size tables,

Table 2
Energy-aware resource allocation mechanisms.

Reference/ Year	Algorithm	Parameters	Improvement	Weakness
Habibi [115] (2017)	QoS-aware routing and energy-efficient VM placement with elephant flow detection (heuristic algorithm).	Energy and SLA.	Improvement of performance and energy.	This research considered an SDN distributed architecture, unlike other approaches.
Jin [120] (2014)	ILP (depth-first, best-fit rule)	Memory capacity and link bandwidth.	Minimize energy.	This method does not consider any VM migration technique.
J. Son [116] (2017)	VM placement and migration algorithm.	Energy and SLA.	Maximize energy cost savings and minimize SLA violation.	This study applied only to homogeneous networks.
Zheng [122] (2014)	Correlation analysis and heuristic algorithm.	Energy.	Less traffic between servers thus saving more power and shorter network delays.	Utilized homogeneous servers and also VM placement does not take into consideration.
Lin [117] (2013)	Dynamic and Disjoint Edge Node Divided Spanning Tree (DENDIST), Traffic-aware flow migration (FM), and Energy-and-Topology-Aware VM Migration (ETA-VMM).	Energy, throughput.	Reducing unwanted traffic in the data center network, unsustainable power consumption due to inadequate routing control, and inappropriate assignment of the VM.	This study only considers the network traffic of the VMs placements to reduce energy consumption.
Liao [124] (2018)	Hybrid partheno-genetic algorithm.	Energy.	A practical solution to the rigid problem in traditional network architecture.	This research does not fully solve the problem of scalability.
Yu [121] (2016)	Heuristics for VM placement and jointly routing.	Energy.	Saves power consumption and improves network performance.	Mechanisms do not care about traffic congestion and ignore memory resources for VMs.
Han, Yoonseon [123] (2015)	Heuristic algorithm for resource raising, a heuristic for resource falling, and heuristic for traffic engineering.	Energy and bandwidth.	Energy reduction.	This method did not consider the lifetime of VDCs and arrived and left time of their requests.

Round Trip Time (RTT) to the controller increases when flow tables are missing and long hosting routes.

Consequently, QVIA-SDN [126] formulates the problem of allocation of physical resources to host VIs on SDN-based data centers in addition to the constrain that appeared when using SDN in the cloud environment. All traditional aspects and SDN problems are expressed as Mixed Integer Program (MIP) and then convert MIP to a linear program and rounding heuristic techniques to reduce internal latency and granted other QoS restrictions when applying QVIA-SDN. It is worth noting that this method did not take VMs placement and migration techniques. Similarly, Cziva et al. [127] presented a server-network framework using SDN based on DC infrastructure to solve the problems resulting from the use of VMs and consolidation in the data centers and the problem that appeared when using SDN in the DCs. Live VM migration is introduced to minimize network-wide connectivity costs and alleviate congestion for higher layers in the DC network hierarchy. Additionally, this proposed architecture can potentially provide connectivity among the network infrastructure and the VMs hosting hypervisors in DC.

Worth noting that some related works were aimed to attain profit provider revenues and reducing consumer cost through jointly virtualized computing and networks. Chase et al. [14] presented a unified algorithm that decides to allocate VMs and network bandwidth in CC based on SDN to reduce consumers' costs. The authors have formulated the problem of stochastic integer programming (SIP) using a two-stage approach to obtain the right decision. This work was only concerned about reserving the resources and was not concerned with the migration and placement process of VMs. Meanwhile, WARM [128] was formulated based on the floodlight controller and cloud controller, aiming to maximize the revenue of cloud providers by considering VMs and network links workloads. WARM schedules a VM and routing path for an application that used a hybrid metaheuristic algorithm called HCSP. Nevertheless, WARM assumes that VMs have static scheduling and ignores their dynamic migration and placement of VMs. A load-balancing algorithm based on Particle Swarm Optimization (PSO) is proposed in [129] to allocate virtual machines and network paths to fulfill cloud user applications in an optimum manner.

The work carried out in [130] focused on utilizing a competition algorithm to select virtual resources that satisfy requests of IaaS as well as employing OpenFlow as an abstraction layer. Performance evaluation when implementing the framework using physical test-bed and Mininet has shown effectiveness in controlling allocating infrastructure by users and reducing request service times and load, which resulted from selecting VMs close to the user. Table 3 shows the different techniques that have adopted the allocation of resources that depend on QoS, the factors that were measured in each study and the algorithm that was applied, and the weaknesses of each technique.

C. Discussion

The joint optimization for solving the power consumption problem of the CDC-based SDN remains a complex problem. Most of the work mentioned above focuses on solving a specific problem, either the migration process for VMs, VMs placement, or consolidation process to reduce the use of both hosts and links. Meanwhile, the joint optimization process must use algorithms to solve all the processes jointly. We find that techniques using routing algorithms focus on routing outside servers and are not concerned with internal links between VMs. Even in the consolidation process, the appropriate number of servers and network bandwidth was not estimated. Hence, meta-heuristic algorithms must be included to improve the joint optimization problems in the future.

The QoS parameters measured within cloud data centers differ, resulting in different problems, limitations, and solutions appearing within cloud computing data centers. Although, the user's QoS requirements are met due to connection and node capacity constraints. However, even in this case, QoS still seems to be NP-hard in joint optimization, which indicates the difficulty of the problem(s). To date, there is no estimation algorithm used to resolve the common optimization technique. Although some approximation algorithms have been suggested (e.g., [14,125]), they simplify the problem and restrict the QoS parameters. As a consequence, only accurate MIP methods, heuristics, and metaheuristics are proposed to address this problem.

Table 3
QoS-aware resource allocation**.**

Reference/Year	Algorithm	Parameters	Improvement	Weakness
de Souza [126] (2019)	Mixed Integer Program (MIP) and rounding heuristic Deterministic Path Search (DPS) algorithm.	Latency, bandwidth.	Heuristic techniques have been used in traffic engineering to reduce data center usage and improve QoS from a consumer perspective. also, the SDN issue was resolved.	This study does not take VMs placement and migration techniques into considerations.
Chase [14] (2014)	Stochastic integer programming.	VM cost and the bandwidth cost.	reserve VMs and network bandwidth to minimize cost.	This method numerical implemented.
Govindarajan [129] (2017)	Particle Swarm Optimization (PSO).	Throughput, resource utilization, performance, and response times.	Achieve QoS.	This study does not directly address their QoS requirements.
Cziva [127] (2016)	Migration algorithm, round-robin, best-fit, and lookahead.	Link utilization, overall communication cost, throughput, VM-to-VM communication cost, and number of migrations.	SDN can manage the network, VMs, and hypervisors.	The major disadvantage of the designed approach was its inability to manage dynamic traffic loads, particularly in congested networks.
Yuan [128] (2018)	A hybrid meta-heuristic algorithm called Hybrid Chaotic Simulated-annealing PSO (HCSP).	RTT.	Increase the profit of cloud providers.	This method focuses on cloud computing only.
J.Son [125] (2019)	Priority-aware VM allocation (PAVA), Bandwidth Allocation (BWA), First-Fit Decreasing (FFD), Dynamic Flow Scheduling Algorithm (DFSA).	Priority, VM capacity, the bandwidth requirement, and energy.	Reduce energy and perform QoS requirement for high priority flows by allocating sufficient bandwidth.	The proposed work considers only the number of cores for defining a VM.
Amarasinghe [130] (2017)	2P-IaaS composition algorithm.	Request service times, efficiency, and scalability.	Full control in allocating IaaS with satisfy QoS.	The major disadvantage in distributed mobile clouds is that service performance can be affected during a user mobility event.

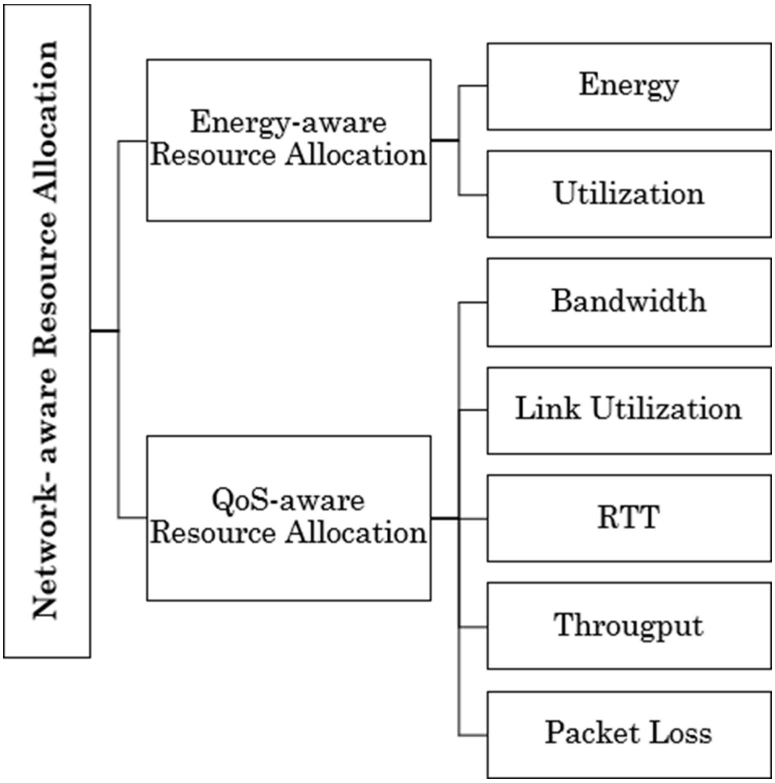

Network- aware Resource Allocation
Energy-aware Resource Allocation
QoS-aware Resource Allocation
Energy
Utilization
Bandwidth
Link Utilization
RTT
Througput
Packet Loss

Fig. 8. Network-aware resource allocation parameters.

Table 4
Energy-aware network resource allocation.

Reference/ Year	Algorithm	Parameters	Improvement	Weakness
Heller [50] (2019)	Greedy bin-packer, topology-aware heuristic, and prediction methods.	Energy	Elastic Tree can make robustness and performance while lowering the energy bill.	This study does not consider VM placement optimization and VM migration.
Subbiah [134] (2016)	Particle Swarm Optimization (PSO) based energy-aware Through open virtual switches.	Energy	Save energy And improve network performance.	The work did not take into account network requirements when applied in the emulator environment.
Xu [136] (2017)	Bandwidth-aware Energy Efficient Routing algorithm with SDN (BEERS).	Link utilization, switch utilization, and energy.	Less energy cost.	There was no guarantee that the network would constrain flows.
Sankari [137] (2016)	VTN Based Energy Efficient Traffic Policy (VTNBEETP) algorithm.	Energy.	Improve scalability and reduce energy.	This research used a specific topology only.

6.2. Network- aware Resource allocation

Many researchers focus on resource allocation problems in the network to address how to provide link bandwidth and routing processes in the switches [16]. In the former network architecture, this problem hinders, especially in cloud computing networks that cause degradation in QoS and power consumption. As the idea of network programming and virtualization appeared, the process of controlling and allocating a network became simplified [131]. In SDN, network virtualization is the method of integrating software and hardware network; hence its functionality performs by a software-based virtual network. When SDN is combined with cloud computing, it facilitates network resource allocation, leading to improved QoS and reduced power. From this perspective, there are two directions for the researchers' a) energy-aware network resource allocation and b) QoS network resource allocation. Fig. 8 shown the parameters used in each case.

A. Energy-aware network resource allocation

The data center's networks consume 10–20% of its total power only [132]. Although this percentage is small relative to the consumption of computing nodes, it reached 3 billion kilowatts in 2006 alone in the U.S [133]. Rising power consumption has limited the potential growth of cloud computing services and has contributed to economic and environmental crises. So, some researchers goal to significantly reduce this rapidly growing energy cost of in-network devices.

Subbiah et al. [134] devised an innovative way to implement the routing algorithm inside switches to detection the best efficient paths that consume less energy to route the packet. Furthermore, An energy-efficient routing algorithm that relies on Particle Swarm Optimization (PSO) has been applied in the Open Daylight Controller [135] to find a routing strategy from source to destination that consumes less power. Also, BEERS [136] is a flow scheduling and routing model based on the SDN controller that allows BEERS to handle SDN switches using the OpenFlow protocol and uses the Northbound APIs to communicate with servers through centralized control. BEERS can detect the routing path by measuring the flow transaction length with flow demands and link utilization. Simulation results demonstrate that BEERS can achieve power saving for the active switches of traffic in the data center. However, BEERS only took the deadline for flow and used only matching links. To demonstrate network scalability in an SDN environment, Subbiah et al. [137] constructed a virtual tenant network utilizing multi-tree rooted topology to simplify the complexity of the network and minimize the expenditure of infrastructure. This work also attains saving power in switches by the proposed VTN Based Energy Efficient Traffic policy algorithm.

ElasticTree [50] is an SDN-based network power manager that dynamically shuts down unused links and switches in the data center to reduce power consumption. ElasticTree is implemented by using three consolidation methods formal model, a greedy bin-packet, topology-aware heuristic. This work can handle unexpected traffic load that happens dynamically in the network. ElasticTree focused on providing power provisioning in network elements but not considered VM placement optimization and VM migration. Studies on energy-aware network resource allocation are shown in Table 4.

B. QoS-aware network resource allocation

In a network, QoS is an essential possession as it is the mechanism that determines how the service is provided. QoS parameters involve assured network bandwidth and latency, loss rate, and congestion control. Despite, SDN Controller offers data flow functionality, customer and cloud provider QoS management remains a tough challenge. In other words, QoS strategy and management techniques that assign constraints of consumer and provider does not provide by SDN Controller. Hence, some studies sought to integrate and develop a mechanism to provide QoS in an SDN controller. As a result, Akella et al. [51] utilizing SDN-based OpenvSwitch [138] to assign QoS bandwidth to the multimedia requirements needed by cloud users. This work serves many cloud consumers with different service applications. This work also used a virtual laboratory to implement this approach.

Govindarajan et al. [52] proposed Q-Ctrl, which slices the network and selects various flow applications to monitor and handle the QoS specifications in the SDN-based cloud infrastructure, the OVS switch, and the Mininet simulator. Their proposed system mainly focuses on small-scale networks to allocate video streaming bandwidth between VMs in the cloud infrastructure. As well as SDN-based network allocation technique is proposed by [139]. Whereas the high-priority jobs have been allocated more bandwidth, depending on the pre-configuration policy taken from the application layer after getting any job weight. This method also extends the floodlight module controller to implement this prototype system.

In order to fulfill the QoS constraints in a multi-path data center operated by an SDN controller, Wang et al. [140] propose a latency-aware flow scheduling approach that meets the QoS requirement and the estimate needed for the bandwidth of the application for each tenant. Therefore, the authors designed MAPLE-Scheduler based on SDN to reschedule some suitable paths to maintain delay performance within QoS objectives while applying load balance across links. In [13], depending on the SDN and the genetic algorithm, an offline network resource pre-allocation model was proposed. Hence, this model can pre-calculate paths between any two virtual machines using the cloud center network and cloud service providers' common multipath feature. Chenhui et al. [141] focus on differentiating the traffic flow into QoS flow and best-effort traffic. The proposed method reroute a feasible high-priority flow path that uses the routing optimization algorithm to ensure specific needs. They incorporate queue technology to assign enough bandwidth for top priority streaming whenever the routing algorithm could not find the appropriate path.

Allocate network resources while reducing total cost in cloud computing based on SDN has been presented by Abdallah et al. [142]. The strategy focuses on allocating VMs based on a centralized SDN controller to avoid the link connection. Notably, this method used licensed software and storage size required on VMs to perform the service conversely, ignoring VMs placement and migration.

Tajiki et al. [143] utilize BLP to formulate the traffic forecast in software-defined cloud networks that both reduce overall bandwidth loss as well as mitigate increased link utilization that is subject to, e.g., delay, bandwidth, and flow protection. Similarly, The work performed at [144] focuses on dynamic network programming with optimized routing traffic in OpenFlow-based networks. This work aims to ensure (QoS) requirements for various applications, proactively prevent resource wastage and congestion, and reduce network burden in the reconfiguration process by applying QoS-aware Network Reconfiguration Relaxed QNR. Table 5 shows various former mechanisms that are used in QoS network resource allocation.

C. Discussion

In data center networks, the power problem is a very important issue, as the question that begs is the appropriate parameters that are measured within the network and the methods used that reduce energy consumption. One of the most popular ways to reduce energy consumption is through consolidation, which has been used to reduce the number of links and switches by leaving unused network devices in sleep or off mode to save power, thus sacrificing network performance, increasing network delay, and unreliable links with higher utilization. Accordingly, much current literature adopts a routing algorithm to overcome energy consumption in network components. The routing algorithm is s still not very satisfied, which encourages us to develop the algorithms and further the efficiency of the algorithms in practice.

As SDN grows, providing QoS to SDN/OpenFlow networks requires more study by research and business. Therefore, the primary objective of QoS is to prioritize QoS parameters, including but not restricted to bandwidth latency and loss of packets. Therefore, the provision of QoS depends mainly on the SLA between side-users and service providers. One such strategy is well adapted to the best effort procedure but does not help manage traffic accurately. We can see how the QoS problem and its parameters have become more difficult to overcome. Noting that most of the work is done concerns only algorithms that solve the bandwidth problem and ignore other parameters, so we need to survey the relevant literature in these areas, considering the new restrictions imposed by its SDN.

6.3. Application-aware resource allocation

Elastic and cost-effective services are provided for the majority of the new Internet applications by using the infrastructure of CC [145]. Hence, handling various requests of the cloud user application is considered a problem in CC, especially are effect totally in QoS in cloud computing. As a result, the application-aware resource allocation in cloud-based SDN has been addressed by researchers who rely on load detection in the application and assign VMs or links as well as various SLAs to the applications. In other words, it describes how applications and resources are delegated to the cloud. Fig. 9 shows the parameters used in the allocation.

App-RA [146] is an OpenFlow-based network resource allocation that uses a neural network to forecast the number of resources that the application needs in a cloud data center and then assign appropriate VMs to satisfy SLA violations and maximize power savings with the help of CICQ switches [147]. Further, to fulfill the network requirements of various applications, App-RS [148] proposed allocating network resources depending on network requirements parameters of each application using the Lagrange Relaxation based aggregated cost Dijkstra algorithm. Likewise, Aziz et al. [90] present modern network provisioning services of application-aware. The fat-tree topology of DCNs is replaced and utilized NEPHELE topology based on the N controller to include hybrid electronic-optical architecture for DCNs.

The research work carried out in [149] focuses o on dynamic QoS

Table 5
QoS-aware network resource allocation.

Reference/Year	Algorithm	Parameters	Improvement	Weakness
Guo-Hong [139] (2017)	Bandwidth allocation strategy.	Bandwidth.	Better performance.	Considering only high priority job.
Govindarajan [52] (2014)	Iperf network monitor tool.	Bandwidth, queue size, and delay.	Applying the QoS in a common environment (SDN, OVS, and Mininet simulator).	Use only default routing technologies. Additionally, this method was only applied to a single scale domain.
Akella [51] (2014)	QoS routing algorithm.	Delay (RTT), available bandwidth, and number of hops.	QoS for all cloud users.	This study does not show Simulation and comparison results.
Wang [140] (2017)	Flow scheduling algorithm.	Link utilization and delay.	Ensuring QoS in multi-pathing data centers.	The case of carrier networks has largely been overlooked.
Guo [13] (2018)	Multi-path genetic algorithm (MPGA) and single-path genetic algorithm.	Bandwidth.	Enhance network resource allocation, gain more provider profit, and satisfy the consumer requirements.	This study only considers the single-path transmission.
Abdelaal [142] (2017)	Network-aware resource allocation strategy.	Link bandwidth.	Reduce the use of upper-layer links.	This study does not take into account the reduced power consumption in a cloud data center environment.
Tajiki [144] (2016)	Binary Linear Programming (BLP).	Maximum link utilization, routing matrix elements, packet loss, and throughput.	QoS requirements for many applications and can proactively prevent congestion and resource waste.	This research does not take into consideration the elephant flows when detecting the congestion.
Tajiki [143] (2017)	QoS-aware Network Reconfiguration (QNR) and relaxed QNR	Network reconfiguration overhead, delay, packet loss.	Effectively reduce restructuring costs according to application QoS flow requirements.	This method ignores the problem of service function chaining (SFC) concerning the energy consumption of the VNFs.
Chenhui [141] (2015)	Lagrange relaxation based aggregated cost and Dijkstra algorithm.	Bandwidth, packet loss, jitter, delay, and throughput.	QoS is assured for all cloud users.	This work used a specific topology only.

Table 6
Application-aware resource allocation.

Reference/ Year	Algorithm	Parameters	Improvement	Weakness
Hong [146] (2014)	Neural network based.	Response time	Meet SLAs, allocate and predictive resources (VMs and network) effectively, reduce application power consumption, and adapt all application types to a cloud data center-based SDN.	This technique focuses only on the response time parameter to check violation of application and predict just VMs resources.
Cheng [148] (2015)	Dijkstra's algorithm and Lagrange relaxation based aggregated cost algorithm.	Tolerance to delay, bandwidth ratio, delay variation, and packet loss rate.	A QoS-conscious routing strategy has been optimized for many groups of applications to satisfy the network requirements of SDN-based cloud data centers.	This mechanism does not solve the SDN problem related to the limiting flow table size of the switch.
Aziz [90] (2017)	REST APIS.	Bandwidth.	Build a new structure of DC-based on SDN.	_
Egilmez [149] (2013)	Lagrangian Relaxation Based Aggregated Cost (LARAC) algorithm.	Delay variation.	Enhance QoS for video application.	This study focuses on one application and takes one SLA parameter.
Cucinotta [150] (2014)	Integer Linear Programming (ILP).	Response times.	Offers high performance in a multi-tenant data center.	This work focuses on cloud computing only.

routing for only video applications that employ the Lagrangian Relaxation Based Aggregated Cost (LARAC) algorithm. Table 6 compares the techniques according to allocate resources in cloud computing based on the application's user's request. Another work focusing on integrated computing and networking resources to optimize performance in the multi-tenant data centers was proposed in [150]. The optimization process studies the application that requires heavy data transfers and tight time constraints and then monitors the state of the underlying physical resources to match the application's requirements.

Papers that consider the allocation of resources as per the needs of each application's needs, which can be accomplished through the current methodology, are the main controls and gimmicks of these algorithms to achieve the specified objectives. Notably, some authors used the LARAC heuristic method to determine the routing optimization problem separately for each source and destination pair, increasing the overall time complexity. Therefore, we need to modify the LARAC algorithm to reduce the number of routing optimization problems. Also, high priority applications must not eliminate the low priority application in the allocation. This allows us further to boost the competitive ratio of algorithms in theory and boost the effectiveness of algorithms in practice by exploring more QoS parameters for different applications.

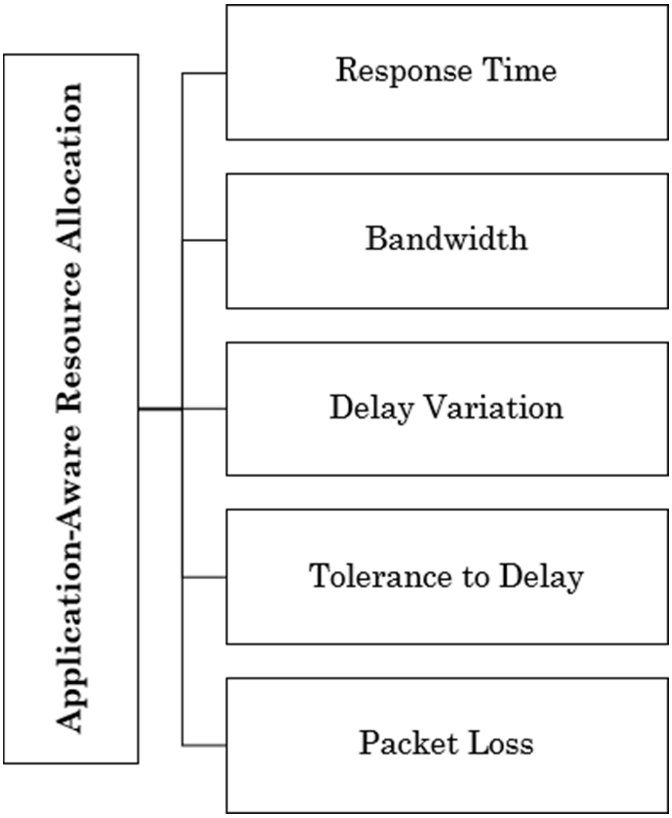


Fig. 9. Application-aware resource allocation parameters.

6.4. Edge -aware resource allocation

Allocate resources in Edge- cloud technology to fulfill the QoS requirement and save power consumption fully is addressed as well as the measured parameters illustrated in Fig. 10.

Workload slicing scheme is built-in [151] to manipulate large data applications in an edge cloud environment. SDN-based control is used to perform the inter-DC migrations and guarantees traffic flow scheduling with power optimization. Also, the Stackelberg game is executed to deliver the best inter-DC migrations. Likewise, MEnSuS [152] design to handle the different incoming workloads from consumers, which can classify types of jobs using an SVM-based scheme. This design reduces SLA violation by using renewable energy sources RES. The switch consolidation strategy has been introduced to save energy usage, delay, and increase bandwidth utilization.

The latency-aware policy was presented by [153], which aims to handle fog traffic steered to DCs. It also offers dynamic resource savings in an optical wide-area SDN that facilitates energy-conscious interaction between cloud and fog. Cao et al. [154] and others suggested a new 5 G IoV architecture based on fog computing and SDN to address the needs of IoV. Whereas the effective use of heterogeneous computing resources to guarantee QoS is a critical problem with this system. Consequently, the authors improved the two architecture algorithms by using the concept of hierarchical clustering to overcome the shortcomings. Noteworthy, experiential results show that the optimized algorithm is capable of obtaining the best experiments. Likewise, IoV setting, the resource allocation scheme derived from other algorithms improves service delay, task execution stability, power consumption, and load balancing.

The research work carried out in [155] introduced a resource allocation framework called IaaSP-SDN to interconnecting the edge cloud data center. IaaSP-SDN used mathematical modeling that implements

Table 7
Edge-aware resource allocation.

Reference/ Year	Algorithm	Parameters	Improvement	Environment
Aujla [152] (2018)	Workload slicing and scheduling algorithm, Energy-aware flow scheduling algorithm, and Stackelberg game for inter-DC migration.	Energy, delay, SLA violations, migration rate, and cost.	Save energy and reduce the delay and cost of the inter-DC relay process.	Edge–cloud computing.
Zaman [155] (2019)	Linearization of quadratic, column generation formulation.	QoS, cost, and path routing.	Provisioning IaaS requests in ECDC, and optimal location to select SDN controller placement.	Edge–cloud computing.
Aujla, [151] (2018)	Workload classification algorithm, server consolidation scheme, and two-stage game for workload scheduling.	Energy, SLA, delay, and bandwidth.	Lesser violations of the SLA, delays, migration costs, and power.	Edge–cloud computing.
Borylo [153] (2016)	Latency Aware policy.	Latency and power.	Reduce latency and carbon footprint.	Fog computing.
Lin [156] (2020)	Edge-cloud SDN (ECSDN) algorithm.	Delay.	Overall performance and QoS of different applications with Various traffic patterns.	———
Cao [154] (2021)	Two-architecture algorithm.	Service delay, stability of task execution, and energy consumption.	Better resource allocation in 5 G IoV.	5 G and fog computing.

IaaS allocating of IaaI requests. Metro Optical Network topology is used to implement the framework under the management of the SDN controller. The authors are also concerned with the position of the SDN controller, and thus they compute several parameters that aid in the process of selecting the position of the SDN controller.

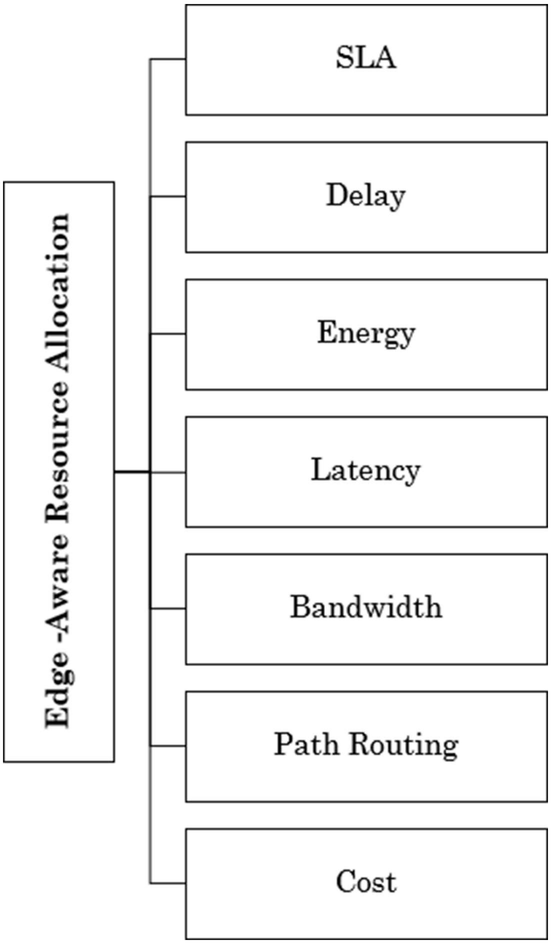


Fig. 10. Edge -aware resource allocation parameters.

Hierarchical edge-cloud SDN (HECSDN) is proposed by [156] to solve the problem of delay resulting from congested by heavy flows related to the limited controller of computation-resource. The suggested model was tested using the MATLAB optimization toolbox, and the results showed the effectiveness of the control system over a large-scale SDN network despite affecting the efficiency and QoS of different network applications. Table 7 illustrate the summary of researchers in edge cloud computing based on SDN.

As in edge clouds, the resource allocation issue must be determined by the location of the user, as well as the wireless network between both the user and the edge server, and the wireless connectivity between the edge server and the cloud server, taking into account the position of the SDN controller. In other words, taking into account the combination of every one of the various QoS parameters and the power consumption will make it very difficult to solve this problem and motivate us to look for it in future work.

7. Open research challenges and future directions

The difficulties of allocating cloud resources revolve around hardware heterogeneity, workload prediction, and requirements of cloud providers and consumers. In this sense, the availability of resources and the optimum use of usable resources for applications to meet the QoS performance objectives in compliance with the SLA is an important issue. In contrast, QoS describes the level of consistency, reliability, and availability provided by the services. Also, it is difficult to delegate due to changing workloads over time that affect various resource requirements of cloud service providers. Likewise, the heterogeneity of the devices and technologies used within the cloud makes resource allocation a challenge. It is worth noting that there are several concerns related to resource allocation in cloud computing systems, including QoS, energy usage, VM migration, provider earnings, utilization expenses, and multi-agent technologies [157].

Even as SDN technology has been implemented into cloud computing, several negatives have arisen. They can be summarized as follows:

1.The size-limited flow table is incorporated in the physical switch.
2.The placement of the SDN controller increases the round-trip time when a flow-table miss occurs.
3.And long host routes.

In the future, we expect the need for modeling components to predict resource requirements, assigning an optimum VM for each application, and choosing the appropriate link to minimize congestion and power consumption. Furthermore, forecasting technology may also be used to predict QoS for each application that interacts dynamically with network boundaries in a cloud-based manner in an effective manner. Worthwhile, most of the methods that were applied to achieve resource allocation in cloud-based on SDN are applied to small-scale experimental and simulation environments, as shown in Fig. 11.

We also need to use meta-heuristic algorithms in the allocation process as they help find the solution in a fast and correct way by combining them with other algorithms that depend on the population, or depend on nature, or rely on biology some exploratory and meta-algorithms based on local search. One of the benefits of integrating two population-based meta-heuristic algorithms would be that the abilities of another algorithm can balance the deficiencies from one algorithm. Furthermore, more research is needed to examine other parameters regardless of power, bandwidth, and predominant usage. The researchers also recommend more research on the following points.

Edge computing: We need to move a massive amount of data across geographically dispersed data centers using backbone networks to efficiently handle big data and the IoT [95,158]. Therefore, the overheads generated by migration between inter-DC cause a high speed of data movements across various DCs and may incur significant costs [159]. Moreover, the positioning of SDN controllers is one of the big issues in the edge cloud data center. Even of many advantages of edge computing, such as location, user mobility, and network connectivity [160], numerous interactions and computing-related issues for future IoT systems now have to be answered [91]. Furthermore, workload forecasting and resource utilization efficiency at both the hardware and software levels require more study.

QoS parameters: Adaptation of SDNs within the cloud computing environment presents new challenges for physical switches and network topology. These points must be considered, including the measurement parameters to ensure the QoS within the cloud computing environment.

Virtual machines allocation: It is the process of placement and migrating a VM. This process is an important issue as it achieves efficient cloud resource scheduling. Although VMs allocations may cause the scarcest bandwidth network resources and congestion resulting from traffic dynamics in the network. Consolidation methods affect the performance and scarcity of resources. Furthermore, flow scheduling mechanisms disregard the specific QoS needed by each VM, which means that they specifically handle each VM's network resources. All these issues pose challenges when allocating VMs and trade-offs to solve them. As most solutions are separately addressed, there must be a mechanism that works to trade-off the solution. Besides, future research needs to be stepped up in this process, which impacts the level of safety risk exposed to VM placement because each VM will have a various level of protection risk [161].

Energy optimization: The energy consumption resulted from network and computing resources is still very large, which negatively affects the total costs of the providers and the users, in addition to environmental pollution [162]. The use of more meta-heuristic algorithms helps to minimize the issue of ensuring the fulfillment of the SLA and energy-saving when taking into account that one is not influenced by the other.

Traffic engineering: The impact of traffic engineering issues in CC based on SDN can be found in [163–166]. These studies only look at network traffic engineering; further research into the traffic or workload of VMs is required.

Resource billing: In cloud computing, it calculates the value of cloud services based on the community and the environment. It is important because the way resources are priced is concerned with distributing limited resources among different cloud users to optimize

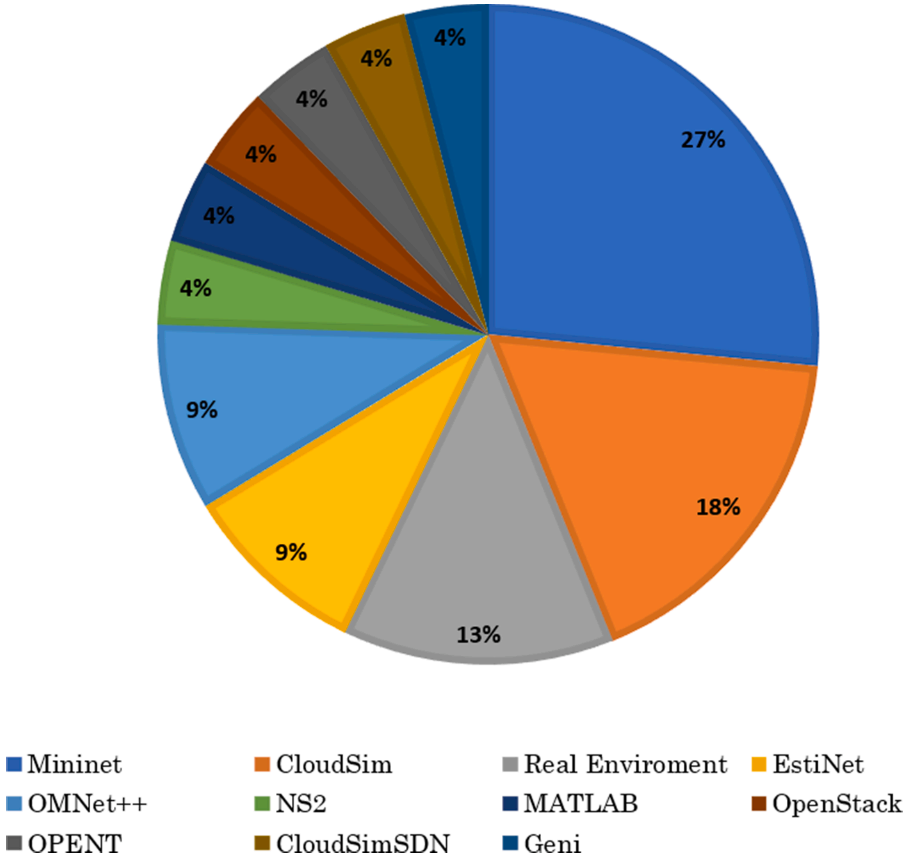


Fig. 11. Types of simulation used in resource allocation.

resource usage. It lowers cloud users' operating expenses while rising cloud providers' benefit and income by optimizing resource usage [167, 127,168,169].

Resource prediction: It is necessary for a collection of workloads running on VMs or PMs to predict the use of computing and network resources (such as CPU, storage, connection bandwidth, and so on) that are required to improve performance. It is also necessary for SLA to calculate the cost of resource use, decide which resource is appropriate for meeting SLA, and evaluate the resources needed [170–173].

Heterogeneous Computing: To provide adequate computing capacity, a modern cloud data center includes many autonomous machines. These machines, on the other hand, may be fitted with a range of devices. Some devices, for instance, have highly powerful GPUs to process artificial intelligence applications, whereas others only have consumer CPUs. In the meantime, network topology might become heterogeneous [174]: Several devices use a gigabit Ethernet network, and others use a wireless mobile network [175] [176]. This inconsistency will obstruct system reliability and resource allocation significantly. As a result, coping with device heterogeneity is a key problem that requires more study for resource allocation mechanisms.

8. Conclusion

In recent years, the effective allocation of resources in the Cloud Data Center (CDC) has emerged as one of the main research issues. This study is aimed to explore and solve the resource allocation concept, which serves as a framework for further research on cloud computing based on SDN in the implementation of resource allocation strategies and to assist future researchers. We investigated resource allocation in Cloud Computing (CC) based on Software-defined Networks (SDN) analysis. We presented new taxonomies based on parameters, algorithms, and optimization techniques based on a comprehensive review of related techniques in literature based on their merits and drawbacks. We addressed the topic of CC in general, along with the related problems and issues that make SDNs to be suitable to be adapted to the CC environment. Moreover, we analyzed the pros and cons of the research allocation of resources mechanisms in two combinations filed and outlined open issues and potential directions.

CRediT authorship contribution statement

Arwa Mohamed: Conceptualization, Data curation, Writing-original draft, Resources. **Mosab Hamdan**: Conceptualization, Data curation, Writing-original draft. **Suleman Khan**: Visualization, Investigation. **Ahmed Abdelaziz***:* Supervision. **Sharief F. Babikir**: Supervision. **MUHAMMAD IMRAN:** Writing-review&editing. **M. N. Marsono:** Writing-review&editing.

Declaration of Competing Interest

The authors declared that there is no conflict of interest among any authors.

References

[1] P. Mell, T. Grance, The NIST definition of cloud computing. National Institute of Standards and Technology, Inf. Technol. Lab. Version 15 (10.07) (2009) 2009.

[2] A.A. Semnanian, J. Pham, B. Englert, X. Wu, Virtualization technology and its impact on computer hardware architecture, in: 2011 Eighth International Conference on Information Technology: New Generations, 2011, pp. 719–724.

[3] N.T. Sultan, D.D. Jamieson, V.A. Simpson, Policy-based forwarding in open shortest path first (OSPF) networks.", Google Patents (Nov. 09, 2010).

[4] C.L. Hedrick, RFC1058: routing information protocol, RFC Editor (1988).

[5] O.N. Fundation, Software-defined networking: the new norm for networks, ONF White Pap 2 (2–6) (2012) 11.

[6] N. McKeown, et al., OpenFlow: enabling innovation in campus networks, ACM SIGCOMM Comput. Commun. Rev. 38 (2) (2008) 69–74.

[7] L. Cui, F.R. Yu, Q. Yan, When big data meets software-defined networking: SDN for big data and big data for SDN, IEEE Netw 30 (1) (2016) 58–65.

[8] R. Mijumbi, J. Serrat, J.-.L. Gorricho, S. Latré, M. Charalambides, D. Lopez, Management and orchestration challenges in network functions virtualization, IEEE Commun. Mag. 54 (1) (2016) 98–105.

[9] W. Ejaz, et al., Internet of Things (IoT) in 5G wireless communications, IEEE Access 4 (2016) 10310–10314.

[10] M. Hamdan, et al., Flow-aware elephant flow detection for software-defined networks, IEEE Access 8 (2020) 72585–72597.

[11] P. Gevros, J. Crowcroft, P. Kirstein, S. Bhatti, Congestion control mechanisms and the best effort service model, IEEE Netw 15 (3) (2001) 16–26.

[12] Z. Wang, J. Crowcroft, Shortest path first with emergency exits, in: Proceedings of the ACM symposium on Communications architectures & protocols, 1990, pp. 166–176.

[13] Y. Guo, Z. Mi, Y. Yang, H. Ma, Efficient Global Network Resource Pre-Allocation in SDN Based Cloud Centers, IEEE Int. Symp. Ind. Electron. 2018-June (2018) 651–656, https://doi.org/10.1109/ISIE.2018.8433617.

[14] J. Chase, R. Kaewpuang, W. Yonggang, D. Niyato, Joint virtual machine and bandwidth allocation in software defined network (SDN) and cloud computing environments, 2014 IEEE Int. Conf. Commun. ICC 2014 (2014) 2969–2974, https://doi.org/10.1109/ICC.2014.6883776.

[15] A.A. Abdelaziz, et al., SDN-based load balancing service for cloud servers, IEEE Commun. Mag. 56 (8) (2018) 106–111, https://doi.org/10.1109/MCOM.2018.1701016.

[16] M. Hamdan, et al., A comprehensive survey of load balancing techniques in software-defined network, J. Netw. Comput. Appl. (2020), 102856, https://doi.org/10.1016/j.jnca.2020.102856.

[17] S.H.H. Madni, M.S. Abd Latiff, Y. Coulibaly, Recent advancements in resource allocation techniques for cloud computing environment: a systematic review, Cluster Comput 20 (3) (2017) 2489–2533.

[18] A. Banumathi and M. Prabakaran, "A SURVEY ON RESOURCE ALLOCATION POLICIES IN CLOUD COMPUTING ENVIRONMENT.".

[19] N. Asha, G.R. Rao, A Review on Various Resource Allocation Strategies in Cloud Computing, Int. J. Emerg. Technol. Adv. Eng. 3 (7) (2013).

[20] V.P. Anuradha, D. Sumathi, A survey on resource allocation strategies in cloud computing, in: International Conference on Information Communication and Embedded Systems (ICICES2014), 2014, pp. 1–7.

[21] N. Hamdy, A. Elsayed, N. ElHaggar, M.-.S. Mostafa, Resource Allocation Strategies in Cloud Computing: overview, Int. J. Comput. Appl. 177 (4) (2017) 18–22, https://doi.org/10.5120/ijca2017915699.

[22] M.H. Mohamaddiah, A. Abdullah, S. Subramaniam, M. Hussin, A survey on resource allocation and monitoring in cloud computing, Int. J. Mach. Learn. Comput. 4 (1) (2014) 31.

[23] B. Jennings, R. Stadler, Resource management in clouds: survey and research challenges, J. Netw. Syst. Manag. 23 (3) (2015) 567–619, https://doi.org/10.1007/s10922-014-9307-7.

[24] A. Yousafzai, et al., Cloud resource allocation schemes: review, taxonomy, and opportunities, Knowl. Inf. Syst. 50 (2) (2017) 347–381, https://doi.org/10.1007/s10115-016-0951-y.

[25] K. Lakhwani, R. Kaur, P. Kumar, M. Thakur, An Extensive Survey on Data Authentication Schemes in Cloud Computing, in: Proc. - 4th Int. Conf. Comput. Sci. ICCS 2018 17, 2019, pp. 59–66, https://doi.org/10.1109/ICCS.2018.00016.

[26] K. Saidi, O. Hioual, and A. Siam, "Resources Allocation in Cloud Computing: a Survey BT - Smart Energy Empowerment in Smart and Resilient Cities," 2020, pp. 356–364.

[27] S. Alnajdi, M. Dogan, E. Al-Qahtani, A Survey on Resource Allocation in Cloud Computing, Int. J. Cloud Comput. Serv. Archit. 6 (5) (2016) 1–11, https://doi.org/10.5121/ijccsa.2016.6501.

[28] K. Wang, Q. Zhou, S. Guo, J. Luo, Cluster frameworks for efficient scheduling and resource allocation in data center networks: a survey, IEEE Commun. Surv. Tutorials 20 (4) (2018) 3560–3580, https://doi.org/10.1109/COMST.2018.2857922.

[29] N. Akhter, M. Othman, Energy aware resource allocation of cloud data center: review and open issues, Cluster Comput 19 (3) (2016) 1163–1182, https://doi.org/10.1007/s10586-016-0579-4.

[30] A. Hameed, et al., A survey and taxonomy on energy efficient resource allocation techniques for cloud computing systems, Computing 98 (7) (2016) 751–774, https://doi.org/10.1007/s00607-014-0407-8.

[31] K. Hamdi, M. Kefi, Network-aware virtual machine placement in cloud data centers: an overview, in: 2016 International Conference on Industrial Informatics and Computer Systems (CIICS), 2016, pp. 1–6.

[32] W. Ejaz, S.K. Sharma, S. Saadat, M. Naeem, A. Anpalagan, and N.A. Chughtai, "A comprehensive survey on resource allocation for CRAN in 5G and beyond networks," *J. Netw. Comput. Appl.*, vol. 160, no. September 2019, 2020, doi: 10.1016/j.jnca.2020.102638.

[33] K.W. Hamaali and S.R.M. Zeebaree, "Resources Allocation for Distributed Systems : a Review," pp. 76–88, 2021, doi: 10.5281/zenodo.4462088.

[34] Y. Xu, G. Gui, S. Member, G. Li, and M. Liu, "A Survey on Resource Allocation for 5G Heterogeneous Networks : current Research, Future Trends and Challenges," no. c, pp. 1–26, 2021, doi: 10.1109/COMST.2021.3059896.

[35] T.Le Duc, R.G. Leiva, P. Casari, P.O. Östberg, Machine learning methods for reliable resource provisioning in edge-cloud computing: a survey, ACM Comput. Surv. 52 (5) (2019), https://doi.org/10.1145/3341145.

[36] R. Su, et al., Resource allocation for network slicing in 5G telecommunication networks: a survey of principles and models, IEEE Netw 33 (6) (2019) 172–179, https://doi.org/10.1109/MNET.2019.1900024.

[37] U. Zehra, M.A. Shah, A survey on resource allocation in software defined networks (SDN), in: ICAC 2017 - 2017 23rd IEEE Int. Conf. Autom. Comput.

Addressing Glob. Challenges through Autom. Comput., 2017, pp. 7–8, https://doi.org/10.23919/IConAC.2017.8082092.

[38] F.A. Nunes De Oliveira -Grr20112021, J. Victor, and T. Risso, "Dynamic Resource Allocation in Software Defined and Virtual Networks: a Comparative Analysis." p. 8, 2014, [Online]. Available: https://pdfs.semanticscholar.org/a3bd/0a309155b02242cd6ee1e887dcf62c61c1ad.pdf.

[39] I. Foster, Y. Zhao, I. Raicu, S. Lu, Cloud computing and grid computing 360-degree compared, 2008 grid computing environments workshop (2008) 1–10.

[40] R. Sherwood, et al., Flowvisor: a network virtualization layer, OpenFlow Switch Consortium, Tech. Rep 1 (2009) 132.

[41] M. Casado, N. Foster, A. Guha, Abstractions for software-defined networks, Commun. ACM 57 (10) (2014) 86–95.

[42] H. Tanaka and T. Koide, "Creating virtual local area network (VLAN)." Google Patents, Feb. 26, 2004.

[43] C.S. Muirhead, D.J. Page, System for supply chain management of virtual private network services. Google Patents, Mar. 23, 2010.

[44] S. Azodolmolky, P. Wieder, R. Yahyapour, SDN-based cloud computing networking, in: Int. Conf. Transparent Opt. Networks, 2013, pp. 1–4, https://doi.org/10.1109/ICTON.2013.6602678.

[45] I. Petri, M. Zou, A.R. Zamani, J. Diaz-Montes, O. Rana, M. Parashar, Integrating software defined networks within a cloud federation, in: 2015 15th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing, 2015, pp. 179–188.

[46] S.-C. Lin, P. Wang, M. Luo, Jointly optimized QoS-aware virtualization and routing in software defined networks, Comput. Networks 96 (2016) 69–78.

[47] A. Chowdhary, S. Pisharody, A. Alshamrani, D. Huang, Dynamic game based security framework in SDN-enabled cloud networking environments, in: Proceedings of the ACM International Workshop on Security in Software Defined Networks & Network Function Virtualization, 2017, pp. 53–58.

[48] H. Amarasinghe, A. Jarray, A. Karmouch, Fault-tolerant IaaS management for networked cloud infrastructure with SDN, in: 2017 IEEE International Conference on Communications (ICC), 2017, pp. 1–7.

[49] K. Zheng, X. Wang, Dynamic control of flow completion time for power efficiency of data center networks, in: 2017 IEEE 37th International Conference on Distributed Computing Systems (ICDCS, 2017, pp. 340–350.

[50] B. Heller, et al., Elastictree: saving energy in data center networks, in: Proc. NSDI 2010 7th USENIX Symp. Networked Syst. Des. Implement., 2019, pp. 249–264.

[51] A.V. Akella, K. Xiong, Quality of service (QoS)-guaranteed network resource allocation via software defined networking (SDN), in: Proceedings - 2014 World Ubiquitous Science Congress: 2014 IEEE 12th International Conference on Dependable, Autonomic and Secure Computing, DASC 2014, 2014, pp. 7–13, https://doi.org/10.1109/DASC.2014.11.

[52] K. Govindarajan, K.C. Meng, H. Ong, W.M. Tat, S. Sivanand, L.S. Leong, Realizing the Quality of Service (QoS) in Software-Defined Networking (SDN) based Cloud infrastructure, in: 2014 2nd Int. Conf. Inf. Commun. Technol. ICoICT 2014, 2014, pp. 505–510, https://doi.org/10.1109/ICoICT.2014.6914113.

[53] D. Clark, J. Rexford, A. Vahdat, A purpose-built global network: google's move to sdn, Commun. ACM 59 (3) (2016) 46–54.

[54] S.S. Manvi, G.K. Shyam, Resource management for Infrastructure as a Service (IaaS) in cloud computing: a survey, J. Netw. Comput. Appl. 41 (2014) 424–440.

[55] M. Armbrust, et al., A view of cloud computing, Commun. ACM 53 (4) (2010) 50–58.

[56] W. Wei, X. Fan, H. Song, X. Fan, J. Yang, Imperfect Information Dynamic Stackelberg Game Based Resource Allocation Using Hidden Markov for Cloud Computing, IEEE Trans. Serv. Comput. 11 (1) (2018) 78–89, https://doi.org/10.1109/TSC.2016.2528246.

[57] Z. Zhou, X. Chen, B. Gu, Multi-Scale Dynamic Allocation of Licensed and Unlicensed Spectrum in Software-Defined HetNets, IEEE Netw 33 (4) (2019) 9–15, https://doi.org/10.1109/MNET.2019.1800469.

[58] J. Liu, J. Wan, B. Zeng, Q. Wang, H. Song, M. Qiu, A scalable and quick-response software defined vehicular network assisted by mobile edge computing, IEEE Commun. Mag. 55 (7) (2017) 94–100, https://doi.org/10.1109/MCOM.2017.1601150.

[59] Z. Zhou, J. Feng, C. Zhang, Z. Chang, Y. Zhang, K.M.S. Huq, SAGECELL: software-defined space-air-ground integrated moving cells, IEEE Commun. Mag. 56 (8) (2018) 92–99, https://doi.org/10.1109/MCOM.2018.1701008.

[60] W. Rafique, L. Qi, I. Yaqoob, M. Imran, R.U. Rasool, W. Dou, Complementing IoT Services through Software Defined Networking and Edge Computing: A Comprehensive Survey, IEEE Commun. Surv. Tutorials 22 (3) (2020) 1761–1804, https://doi.org/10.1109/COMST.2020.2997475.

[61] K. Xu, X. Wang, W. Wei, H. Song, B. Mao, Toward software defined smart home, IEEE Commun. Mag. 54 (5) (2016) 116–122, https://doi.org/10.1109/MCOM.2016.7470945.

[62] M. Awais, A. Ahmed, S.A. Ali, M. Naeem, W. Ejaz, A. Anpalagan, Resource management in multicloud IoT radio access network, IEEE Internet Things J 6 (2) (2019) 3014–3023, https://doi.org/10.1109/JIOT.2018.2878511.

[63] R. Mikkilineni, V. Sarathy, Cloud Computing and the Lessons from the Past. 2009 18th IEEE International Workshops On Enabling Technologies: Infrastructures for Collaborative Enterprises, 2009, pp. 57–62.

[64] M. Sharkh, M. Jammal, A. Shami, A. Ouda, Resource allocation in a network-based cloud computing environment: design challenges, IEEE Commun. Mag. 51 (11) (2013) 46–52, https://doi.org/10.1109/MCOM.2013.6658651.

[65] S. Iqbal, et al., On cloud security attacks: a taxonomy and intrusion detection and prevention as a service, J. Netw. Comput. Appl. 74 (2016) 98–120.

[66] I. Pietri, R. Sakellariou, Mapping virtual machines onto physical machines in cloud computing: a survey, ACM Comput. Surv. 49 (3) (2016), https://doi.org/10.1145/2983575.

[67] Y.-J. Wang, W.-D. Sun, S. Zhou, X.-Q. Pei, X.-Y. Li, Key technologies of distributed storage for cloud computing, Ruanjian Xuebao/Journal Softw 23 (4) (2012) 962–986.

[68] S.P. Crago, J.P. Walters, Heterogeneous cloud computing: the way forward, Computer (Long. Beach. Calif). 48 (1) (2015) 59–61.

[69] K. Ramachandran, R. Kokku, R. Mahindra, S. Rangarajan, N. Brunswick, 60GHz Data-Center Networking: wireless->Worry less? Spectrum 1 (2008) 1–11.

[70] S. Ilager, K. Ramamohanarao, R. Buyya, ETAS: energy and thermal-aware dynamic virtual machine consolidation in cloud data center with proactive hotspot mitigation, Concurr. Comput. 31 (17) (2019) 1–15, https://doi.org/10.1002/cpe.5221.

[71] J. Singh, Study of response time in cloud computing, Int. J. Inf. Eng. Electron. Bus. 6 (5) (2014) 36.

[72] E. Bauer, R. Adams, Reliability and Availability of Cloud Computing, John Wiley & Sons, 2012.

[73] A. Iosup, S. Ostermann, M.N. Yigitbasi, R. Prodan, T. Fahringer, D. Epema, Performance analysis of cloud computing services for many-tasks scientific computing, IEEE Trans. Parallel Distrib. Syst. 22 (6) (2011) 931–945.

[74] S.H.H. Madni, M.S. Abd Latiff, Y. Coulibaly, Resource scheduling for infrastructure as a service (IaaS) in cloud computing: challenges and opportunities, J. Netw. Comput. Appl. 68 (2016) 173–200.

[75] A.K. Baughman, L.M. Boyer, C.F. Codella, R.L. Darden, W.G. Dubyak, A. Greenland, Workload adaptive cloud computing resource allocation. Google Patents, Jul. 29, 2014.

[76] H.K. Ala'a Al-Shaikh, A. Sharieh, A. Sleit, Resource utilization in cloud computing as an optimization problem, Resource 7 (6) (2016).

[77] S. Mustafa, B. Nazir, A. Hayat, S.A. Madani, Resource management in cloud computing: taxonomy, prospects, and challenges, Comput. Electr. Eng. 47 (2015) 186–203.

[78] P. Patel, A.H. Ranabahu, and A.P. Sheth, "Service level agreement in cloud computing," 2009.

[79] A. Beloglazov, J. Abawajy, R. Buyya, Energy-aware resource allocation heuristics for efficient management of data centers for cloud computing, Futur. Gener. Comput. Syst. 28 (5) (2012) 755–768.

[80] S. Chaisiri, B.-S. Lee, D. Niyato, Optimization of resource provisioning cost in cloud computing, IEEE Trans. Serv. Comput. 5 (2) (2011) 164–177.

[81] Y. Zhang, Network bandwidth allocation in multi-tenancy cloud computing networks. Google Patents, Jun. 23, 2015.

[82] T.-C. Yen, C.-S. Su, An SDN-based cloud computing architecture and its mathematical model, 2014 International Conference on Information Science, Electronics and Electrical Engineering 3 (2014) 1728–1731.

[83] H. Li, K. Ota, M. Dong, Virtual network recognition and optimization in SDN-enabled cloud environment, IEEE Trans. Cloud Comput. (2018).

[84] F. Li, J. Cao, X. Wang, Y. Sun, Y. Sahni, Enabling software defined networking with qos guarantee for cloud applications, in: *2017 IEEE 10th International Conference on Cloud Computing (CLOUD)*, 2017, pp. 130–137.

[85] B. Jiang, Q. He, X. Li, H. Huang, QoS Control Method Based on SDN for Mobile Cloud Service, in: 2020 IEEE 13th International Conference on Cloud Computing (CLOUD), 2020, pp. 275–283.

[86] A.J. Kadhim, S.A.H. Seno, Maximizing the utilization of fog computing in internet of vehicle using sdn, IEEE Commun. Lett. 23 (1) (2018) 140–143.

[87] S. Al-Mashhadi, M. Anbar, R.A. Jalal, A. Al-Ani, Design of cloud computing load balance system based on SDN technology. Computational Science and Technology, Springer, 2020, pp. 123–133.

[88] S.K. Mishra, S. Sahoo, B. Sahoo, S.K. Jena, Energy-Efficient Service Allocation Techniques in Cloud: a Survey, IETE Tech. Rev. (Institution Electron. Telecommun. Eng. India) 0 (0) (2019) 1–14, https://doi.org/10.1080/02564602.2019.1620648.

[89] V. Persico, A. Botta, A. Montieri, A. Pescapé, A first look at public-cloud inter-datacenter network performance, in: 2016 IEEE Glob. Commun. Conf. GLOBECOM 2016 - Proc, 2016, https://doi.org/10.1109/GLOCOM.2016.7841498.

[90] M. Aziz, H.A. Fazely, G. Landi, D. Gallico, K. Christodoulopoulos, P. Wieder, SDN-enabled application-aware networking for data center networks, in: 2016 IEEE Int. Conf. Electron. Circuits Syst. ICECS 2016, 2017, pp. 372–375, https://doi.org/10.1109/ICECS.2016.7841210.

[91] A. Alnoman, S.K. Sharma, W. Ejaz, A. Anpalagan, Emerging edge computing technologies for distributed IoT systems, IEEE Netw 33 (6) (2019) 140–147.

[92] X. Wang, M. Chen, T. Taleb, A. Ksentini, V.C.M. Leung, Cache in the air: exploiting content caching and delivery techniques for 5G systems, IEEE Commun. Mag. 52 (2) (2014) 131–139.

[93] D. Liu, B. Chen, C. Yang, A.F. Molisch, Caching at the wireless edge: design aspects, challenges, and future directions, IEEE Commun. Mag. 54 (9) (2016) 22–28.

[94] E. Ahmed, A. Naveed, A. Gani, S.H. Ab Hamid, M. Imran, M. Guizani, Process state synchronization-based application execution management for mobile edge/cloud computing, Futur. Gener. Comput. Syst. 91 (2019) 579–589.

[95] E. Ahmed, et al., Bringing computation closer toward the user network: is edge computing the solution? IEEE Commun. Mag. 55 (11) (2017) 138–144.

[96] F. Bonomi, R. Milito, J. Zhu, S. Addepalli, Fog computing and its role in the internet of things, in: Proceedings of the first edition of the MCC workshop on Mobile cloud computing, 2012, pp. 13–16.

[97] IEEE Communications Society, Internet of Things Emerging Technologies Initiatives. IEEE Computational Intelligence Society, Institute of Electrical and Electronics Engineers, and S. Internet of Things Week (2017 : Geneva, "GIoTS2017 : Global Internet of Things Summit : 2017 Proceedings papers : CICG, Geneva, June 6-9, 2017.,", 2017.

[98] X. Li, D. Li, J. Wan, C. Liu, M. Imran, Adaptive transmission optimization in SDN-based industrial Internet of Things with edge computing, IEEE Internet Things J 5 (3) (2018) 1351–1360.

[99] J. Son, A.V. Dastjerdi, R.N. Calheiros, X. Ji, Y. Yoon, R. Buyya, CloudSimSDN: modeling and simulation of software-defined cloud data centers, in: Proc. - 2015 IEEE/ACM 15th Int. Symp. Clust. Cloud, Grid Comput. CCGrid 2015, 2015, pp. 475–484, https://doi.org/10.1109/CCGrid.2015.87.

[100] "Miminet." www.mininet.org.

[101] "GitHub - noxrepo/pox: the POX network software platform." https://github.com/noxrepo/pox (accessed Mar. 25, 2021).

[102] J. Teixeira, G. Antichi, D. Adami, A. Del Chiaro, S. Giordano, A. Santos, Datacenter in a box: test your SDN cloud-datacenter controller at home. *2013 S European Workshop On Software Defined Networks*, 2013, pp. 99–104.

[103] "OpenStack Foundation, 'Open source software for creating private and public clouds.'" https://www.openstack.org/.

[104] C.H. Benet, R. Nasim, K.A. Noghani, A. Kassler, OpenStackEmu—A cloud testbed combining network emulation with OpenStack and SDN, in: 2017 14th IEEE Annual Consumer Communications & Networking Conference (CCNC), 2017, pp. 566–568.

[105] S.H.H. Madni, M.S.A. Latiff, Y. Coulibaly, S.M. Abdulhamid, An appraisal of meta-heuristic resource allocation techniques for IaaS cloud, Indian J. Sci. Technol. 9 (4) (2016) 1–14.

[106] X.-.S. Yang, Nature-inspired Metaheuristic Algorithms, Luniver press, 2010.

[107] S. Goudarzi, M.H. Anisi, H. Ahmadi, L. Musavian, Dynamic Resource Allocation Model for Distribution Operations using SDN, IEEE Internet Things J (2020).

[108] K.T. Bagci, A.M. Tekalp, Dynamic resource allocation by batch optimization for value-added video services over SDN, IEEE Trans. Multimed. 20 (11) (2018) 3084–3096.

[109] P. Sjövall, A. Oinonen, M. Teuho, J. Vanne, T.D. Hämäläinen, Dynamic resource allocation for HEVC encoding in FPGA-accelerated SDN cloud, in: 2019 IEEE Nordic Circuits and Systems Conference (NORCAS): NORCHIP and International Symposium of System-on-Chip (SoC), 2019, pp. 1–5.

[110] B. Chen, Y. Zhang, G. Iosifidis, M. Liu, Reinforcement Learning on Computational Resource Allocation of Cloud-based Wireless Networks. 2020 IEEE 6th World Forum On Internet of Things (WF-IoT), 2020, pp. 1–6.

[111] F.-.H. Tseng, X. Wang, L.-.D. Chou, H.-.C. Chao, V.C.M. Leung, Dynamic resource prediction and allocation for cloud data center using the multiobjective genetic algorithm, IEEE Syst. J. 12 (2) (2017) 1688–1699.

[112] A. Martin, et al., Network resource allocation system for QoE-aware delivery of media services in 5G networks, IEEE Trans. Broadcast. 64 (2) (2018) 561–574.

[113] J. Son, R. Buyya, A taxonomy of software-defined networking (SDN)-enabled cloud computing, ACM Comput. Surv. 51 (3) (2018), https://doi.org/10.1145/3190617.

[114] Z. Zhou, J. Gong, Y. He, Y. Zhang, Software defined machine-to-machine communication for smart energy management, IEEE Commun. Mag. 55 (10) (2017) 52–60.

[115] P. Habibi, M. Mokhtari, M. Sabaei, QRVE: qoS-aware routing and energy-efficient VM Placement for Software-Defined DataCenter Networks, in: 2016 8th Int. Symp. Telecommun. IST 2016, 2017, pp. 533–539, https://doi.org/10.1109/ISTEL.2016.7881879.

[116] J. Son, A.V. Dastjerdi, R.N. Calheiros, R. Buyya, SLA-Aware and energy-efficient dynamic overbooking in SDN-based cloud data centers, IEEE Trans. Sustain. Comput. 2 (2) (2017) 76–89, https://doi.org/10.1109/TSUSC.2017.2702164.

[117] W.C. Lin, C.H. Liao, K.T. Kuo, C.H.P. Wen, Flow-and-VM migration for optimizing throughput and energy in SDN-based cloud datacenter, Proc. Int. Conf. Cloud Comput. Technol. Sci. CloudCom 1 (2013) 206–211, https://doi.org/10.1109/CloudCom.2013.35.

[118] "The Network Simulator - ns-2." https://www.isi.edu/nsnam/ns/(accessed May 20, 2020).

[119] "The CLOUDS Lab: Flagship Projects - Gridbus and Cloudbus." http://www.cloudbus.org/cloudsim/(accessed May 20, 2020).

[120] H. Jin, et al., Joint host-network optimization for energy-efficient data center networking, in: Proc. - IEEE 27th Int. Parallel Distrib. Process. Symp. IPDPS 2013, 2013, pp. 623–634, https://doi.org/10.1109/IPDPS.2013.100.

[121] B. Yu, Y. Han, X. Wen, X. Chen, Z. Xu, An energy-aware algorithm for optimizing resource allocation in software defined network, in: 2016 IEEE Global Communications Conference, GLOBECOM 2016 - Proceedings, 2016, pp. 1–7, https://doi.org/10.1109/GLOCOM.2016.7841589.

[122] K. Zheng, X. Wang, L. Li, X. Wang, Joint power optimization of data center network and servers with correlation analysis, in: Proc. - IEEE INFOCOM, 2014, pp. 2598–2606, https://doi.org/10.1109/INFOCOM.2014.6848207.

[123] Y. Han, J. Li, J.Y. Chung, J.H. Yoo, J.W.K. Hong, SAVE: energy-Aware Virtual Data Center embedding and Traffic Engineering using SDN, in: 1st IEEE Conf. Netw. Softwarization Software-Defined Infrastructures Networks, Clouds, IoT Serv. NETSOFT 2015, 2015, https://doi.org/10.1109/NETSOFT.2015.7116142. Vm.

[124] Q. Liao, Z. Wang, Energy consumption optimization scheme of cloud data center based on SDN, Procedia Comput. Sci. 131 (2018) 1318–1327, https://doi.org/10.1016/j.procs.2018.04.327.

[125] J. Son, R. Buyya, Priority-Aware VM Allocation and Network Bandwidth Provisioning in Software-Defined Networking (SDN)-Enabled Clouds, IEEE Trans. Sustain. Comput. 4 (1) (2019) 17–28, https://doi.org/10.1109/TSUSC.2018.2842074.

[126] F.R. de Souza, C.C. Miers, A. Fiorese, M.D. de Assunção, G.P. Koslovski, QVIA-SDN: towards QoS-Aware Virtual Infrastructure Allocation on SDN-based Clouds, J. Grid Comput. 17 (3) (2019) 447–472, https://doi.org/10.1007/s10723-019-09479-x.

[127] R. Cziva, S. Jouët, D. Stapleton, F.P. Tso, D.P. Pezaros, SDN-Based Virtual Machine Management for Cloud Data Centers, IEEE Trans. Netw. Serv. Manag. 13 (2) (2016) 212–225, https://doi.org/10.1109/TNSM.2016.2528220.

[128] H. Yuan, J. Bi, M. Zhou, K. Sedraoui, WARM: workload-Aware Multi-Application Task Scheduling for Revenue Maximization in SDN-Based Cloud Data Center, IEEE Access 6 (2018) 645–657, https://doi.org/10.1109/ACCESS.2017.2773645.

[129] K. Govindarajan, V.S. Kumar, An intelligent load balancer for software defined networking (SDN) based cloud infrastructure, in: Proc. 2017 2nd IEEE Int. Conf. Electr. Comput. Commun. Technol. ICECCT 2017, 2017, https://doi.org/10.1109/ICECCT.2017.8117881.

[130] H. Amarasinghe, A. Karmouch, SDN-based framework for infrastructure as a service clouds. IEEE Int. Conf. Cloud Comput. CLOUD, 2017, pp. 782–789, https://doi.org/10.1109/CLOUD.2016.106.

[131] S. Khan, et al., Software-defined network forensics: motivation, potential locations, requirements, and challenges, IEEE Netw 30 (6) (2016) 6–13.

[132] A. Greenberg, J. Hamilton, D.A. Maltz, P. Patel, The Cost of a cloud: Research Problems in Data Center Networks, ACM New York, NY, USA, 2008.

[133] E. Energy Star, "Report to Congress on Server and Data Center Energy Efficiency Public Law 109-431," 2007.

[134] S. Subbiah, V. Perumal, Energy-aware network resource allocation in SDN, in: Proc. 2016 IEEE Int. Conf. Wirel. Commun. Signal Process. Networking, WiSPNET 2016, 2016, pp. 2071–2075, https://doi.org/10.1109/WiSPNET.2016.7566506.

[135] "opendaylight, " 2018, [Online]. Available: https://www.opendaylight.org/.

[136] G. Xu, B. Dai, B. Huang, J. Yang, S. Wen, Bandwidth-aware energy efficient flow scheduling with SDN in data center networks, Futur. Gener. Comput. Syst. 68 (2017) 163–174, https://doi.org/10.1016/j.future.2016.08.024.

[137] S. Sankari, V. Perumal, Network resource provisioning in cloud data center across manifold SDN controllers, in: Proc. 2nd IEEE Int. Conf. Eng. Technol. ICETECH 2016, 2016, pp. 543–548, https://doi.org/10.1109/ICETECH.2016.7569311. March.

[138] [*Online*], ""Openvswitch."," *Available*, [Online]. Available: http://www.openvswitch.org/.

[139] Z. Guo-Hong, Network resource scheduling mechanism of cloud computing based on SDN, in: Proc. - 2016 Int. Conf. Intell. Transp. Big Data Smart City, ICITBS 2016, 2017, pp. 332–336, https://doi.org/10.1109/ICITBS.2016.60.

[140] R. Wang, S. Mangiante, A. Davy, L. Shi, B. Jennings, QoS-aware multipathing in datacenters using effective bandwidth estimation and SDN, in: 2016 12th Int. Conf. Netw. Serv. Manag. CNSM 2016 Work. 3rd Int. Work. Manag. SDN NFV, ManSDN/NFV 2016, Int. Work. Green ICT Smart Networking, GISN 2016, 2017, pp. 342–347, https://doi.org/10.1109/CNSM.2016.7818444.

[141] C. Xu, B. Chen, P. Fu, H. Qian, A dynamic resource allocation model for guaranteeing quality of service in software defined networking based cloud computing environment, in: International Conference on Cloud Computing and Security, 2015, pp. 206–217.

[142] M.A. Abdelaal, G.A. Ebrahim, W.R. Anis, Network-aware resource management strategy in cloud computing environments, in: Proc. 2016 11th Int. Conf. Comput. Eng. Syst. ICCES 2016, 2017, pp. 26–31, https://doi.org/10.1109/ICCES.2016.7821970.

[143] M.M. Tajiki, B. Akbari, N. Mokari, Optimal Qos-aware network reconfiguration in software defined cloud data centers, Comput. Networks 120 (2017) 71–86, https://doi.org/10.1016/j.comnet.2017.04.003.

[144] I. Symposium, "QRTP : qoS-Aware Resource Reallocation Based on Traffic Prediction in Software Defined Cloud Networks," pp. 527–532, 2016.

[145] R. Buyya, C.S. Yeo, S. Venugopal, J. Broberg, I. Brandic, Cloud computing and emerging IT platforms: vision, hype, and reality for delivering computing as the 5th utility, Futur. Gener. Comput. Syst. 25 (6) (2009) 599–616.

[146] W. Hong, K. Wang, Y.H. Hsu, Application-aware resource allocation for SDN-based cloud datacenters, in: Proc. - 2013 Int. Conf. Cloud Comput. Big Data, CLOUDCOM-ASIA 2013, 2013, pp. 106–110, https://doi.org/10.1109/CLOUDCOM-ASIA.2013.44.

[147] H. Jin, D. Pan, J. Liu, N. Pissinou, OpenFlow-based flow-level bandwidth provisioning for CICQ switches, IEEE Trans. Comput. 62 (9) (2012) 1799–1812.

[148] L.C. Cheng, K. Wang, Y.H. Hsu, Application-aware Routing Scheme for SDN-based cloud datacenters, in: International Conference on Ubiquitous and Future Networks, ICUFN, 2015, pp. 820–825, https://doi.org/10.1109/ICUFN.2015.7182657, 2015-Augus.

[149] H.E. Egilmez, S. Civanlar, A.M. Tekalp, An optimization framework for QoS-enabled adaptive video streaming over openflow networks, IEEE Trans. Multimed. 15 (3) (2013) 710–715, https://doi.org/10.1109/TMM.2012.2232645.

[150] T. Cucinotta, D. Lugones, D. Cherubini, E. Jul, Data centre optimisation enhanced by software defined networking, in: IEEE Int. Conf. Cloud Comput. CLOUD, 2014, pp. 136–143, https://doi.org/10.1109/CLOUD.2014.28.

[151] G.S. Aujla, N. Kumar, A.Y. Zomaya, R. Ranjan, Optimal decision making for big data processing at edge-cloud environment: an SDN perspective, IEEE Trans. Ind. Informatics 14 (2) (2018) 778–789, https://doi.org/10.1109/TII.2017.2738841.

[152] G.S. Aujla, N. Kumar, MEnSuS: an efficient scheme for energy management with sustainability of cloud data centers in edge–cloud environment, Futur. Gener. Comput. Syst. 86 (2018) 1279–1300, https://doi.org/10.1016/j.future.2017.09.066.

[153] P. Borylo, A. Lason, J. Rzasa, A. Szymanski, A. Jajszczyk, Energy-aware fog and cloud interplay supported by wide area software defined networking, in: 2016 IEEE Int. Conf. Commun. ICC 2016, 2016, https://doi.org/10.1109/ICC.2016.7511451.

[154] B. Cao, Z. Sun, J. Zhang, Y. Gu, Resource Allocation in 5G IoV Architecture Based on SDN and Fog-Cloud Computing, IEEE Trans. Intell. Transp. Syst. (2021) 1–9, https://doi.org/10.1109/TITS.2020.3048844.

[155] F.A. Zaman, A. Jarray, A. Karmouch, Software Defined Network-Based Edge Cloud Resource Allocation Framework, IEEE Access 7 (January) (2019) 10672–10690, https://doi.org/10.1109/ACCESS.2018.2889943.

[156] F.P.C. Lin, Z. Tsai, Hierarchical Edge-Cloud SDN Controller System with Optimal Adaptive Resource Allocation for Load-Balancing, IEEE Syst. J. 14 (1) (2020) 265–276, https://doi.org/10.1109/JSYST.2019.2894689.

[157] E. Alkayal, "Optimizing Resource Allocation using Multi-Objective Particle Swarm Optimization in Cloud Computing Systems," no. January 2018.

[158] B. Mohammed, et al., Edge Computing Intelligence Using Robust Feature Selection for Network Traffic Classification in Internet-of-Things, IEEE Access 8 (2020) 224059–224070.

[159] A. Yousafzai, I. Yaqoob, M. Imran, A. Gani, R.M. Noor, Process migration-based computational offloading framework for IoT-supported mobile edge/cloud computing, IEEE Internet Things J 7 (5) (2019) 4171–4182.

[160] A. Gani, G.M. Nayeem, M. Shiraz, M. Sookhak, M. Whaiduzzaman, S. Khan, A review on interworking and mobility techniques for seamless connectivity in mobile cloud computing, J. Netw. Comput. Appl. 43 (2014) 84–102.

[161] J. Anderson, "Software Defined Network Based Virtual Machine Placement in Cloud Systems," pp. 876–881, 2017.

[162] S. Shrabanee, A.K. Rath, SDN-cloud: a power aware resource management system for efficient energy optimization, Int. J. Intell. Unmanned Syst. (2020).

[163] M. Bastam, M. Sabaei, R. Yousefpour, A scalable traffic engineering technique in an SDN-based data center network, Trans. Emerg. Telecommun. Technol. 29 (2) (2018) 1–16, https://doi.org/10.1002/ett.3268.

[164] I.F. Akyildiz, A. Lee, P. Wang, M. Luo, W. Chou, A roadmap for traffic engineering in SDN-OpenFlow networks, Comput. Networks 71 (2014) 1–30.

[165] H. Huang, S. Guo, P. Li, B. Ye, I. Stojmenovic, Joint optimization of rule placement and traffic engineering for QoS provisioning in software defined network, IEEE Trans. Comput. 64 (12) (2015) 3488–3499.

[166] H. Huang, S. Guo, J. Wu, J. Li, Green datapath for TCAM-based software-defined networks, IEEE Commun. Mag. 54 (11) (2016) 194–201.

[167] B. Yi, X. Wang, M. Huang, Y. Zhao, Novel resource allocation mechanism for SDN-based data center networks, J. Netw. Comput. Appl. 155 (September 2019) 2020, https://doi.org/10.1016/j.jnca.2020.102554.

[168] H. Eghbali, V.W.S. Wong, Bandwidth allocation and pricing for SDN-enabled home networks, in: 2015 IEEe international conference on communications (ICC), 2015, pp. 5342–5347.

[169] Y. Gu, J. Tao, X. Wu, X. Ma, Online mechanism with latest-reservation for dynamic VMs allocation in private cloud, Int. J. Syst. Assur. Eng. Manag. 8 (3) (2017) 2009–2016.

[170] M.M. Tajiki, B. Akbari, M. Shojafar, N. Mokari, Joint QoS and congestion control based on traffic prediction in SDN, Appl. Sci. 7 (12) (2017) 1265.

[171] B. Martini, et al., An SDN orchestrator for resources chaining in cloud data centers, in: 2014 European Conference on Networks and Communications (EuCNC), 2014, pp. 1–5.

[172] M. Gharbaoui, B. Martini, D. Adami, S. Giordano, P. Castoldi, Cloud and network orchestration in SDN data centers: design principles and performance evaluation, Comput. Networks 108 (2016) 279–295, https://doi.org/10.1016/j.comnet.2016.08.029.

[173] A. Mohamed, M. Hamdan, A. Abdelazizb, S.F. Babiker, Dynamic Resource Allocation in Cloud Computing Based on Software-Defined Networking Framework, Open J. Sci. Technol. 3 (3) (2020) 304–313.

[174] J. An, K. Yang, J. Wu, N. Ye, S. Guo, Z. Liao, Achieving sustainable ultra-dense heterogeneous networks for 5G, IEEE Commun. Mag. 55 (12) (2017) 84–90.

[175] R. Atat, L. Liu, H. Chen, J. Wu, H. Li, Y. Yi, Enabling cyber-physical communication in 5G cellular networks: challenges, spatial spectrum sensing, and cyber-security, IET Cyber-Physical Syst. Theory Appl. 2 (1) (2017) 49–54.

[176] K. Wang, Y. Wang, D. Zeng, S. Guo, An SDN-based architecture for next-generation wireless networks, IEEE Wirel. Commun. 24 (1) (2017) 25–31.

Arwa Mohamed received the B.Sc. degree in computer and electronic systems engineering from the University of Science and Technology, Sudan, in 2008, and the M.Sc. degree in computer architecture and networking from the university of Khartoum, Sudan, in 2014. She is currently pursuing the Ph.D. degree with the Faculty of Electrical and Electronic Engineering, University of Khartoum. Her current research interests are in software defined networking (SDN), cloud computing, resources allocation, and future networks.

Mosab Hamdan received the B.Sc. degree in computer and Electronic System Engineering from the University of Science and Technology (UST), Sudan, in 2010, the M.Sc. degree in Computer Architecture and Networking from the University of Khartoum (UofK), Sudan, in 2014, and the Ph.D. degree in Electrical Engineering (Computer Networking) from the Faculty of Engineering, School of Electrical Engineering, Universiti Teknologi Malaysia (UTM), Malaysia, in 2021. From 2010 to 2015, he was a teaching assistant and lecturer with the Department of Computer and Electronics System Engineering, Faculty of Engineering, University of Science and Technology (UST). He is currently a Researcher with the Universiti Teknologi Malaysia under the Post-Doctoral Fellowship Scheme. His current research interests are software-defined networking (SDN), load balancing, network traffic classification, internet-of-things (IoT), cloud computing, network security, and future network.

Suleman Khan received the Ph.D. degree (Distinction) in computer science and information technology from the Universiti Malaya, Malaysia, in 2017. He was a Faculty Member of the School of Information Technology, Monash University, Malaysia, from June 2017 to March 2019. He is currently a Faculty Member of the Department of Computer and Information Sciences, Northumbria University, Newcastle upon Tyne, U.K. He has published more than 60 high-impact research articles in reputed international journals and conferences. His research areas include, but are not limited to, network forensics, software-defined networks, the Internet-of-Things, cloud computing, and vehicular communications.

Ahmed Abdelaziz received the M.*Sc*. degree in computer science and the Ph.D. degree in information technology from the Universiti Malaya (UM), Malaysia, in 2007 and 2017, respectively. He has been working on ONOS and Open Stack, since October 2015, during the Ph.D. degree research project. In the Ph.D. degree research, he proposed a novel service-based load balancing technique to use in the cloud using SDN and OpenStack. He is currently a full-time Assistant Professor with Future University (FU), Sudan. He published a number of ISI index articles in the areas of SDN, OpenFlow, and network virtualization. He has been involved in the centre for Mobile Cloud Computing Research (C4MCCR) Projects funded by the Malaysian Ministry of Higher Education. His areas of interest include SDN/NFV technology, OpenStack, and network virtualization.

Sharief F. Babikir is a professor of Electronics. He studied Electrical Engineering at the University of Khartoum and obtained his PhD from the University of Glasgow. He worked in the United Kingdom in the academic and the industrial sectors. Currently he is the General Director of Africa City of Technology and he is a Technical consultant with Sudan Electricity Distribution Company. He is a Senior Member of the IEEE and is the Chairman of IEEE Sudan Subsection.

Muhammad Imran is an Associate Professor in the College of Applied Computer Science at King Saud University, Saudi Arabia. He received a PhD in Information Technology from the University Teknologi PETRONAS, Malaysia in 2011. His research interest includes Internet of Things, Mobile and Wireless Networks, Big Data Analytics, Cloud computing, and Information Security. His research is financially supported by several grants. He has completed a number of international collaborative research projects with reputable universities. He has published more than 250 research articles in peer-reviewed, well-recognized international conferences and journals. Many of his research articles are among the highly cited and most downloaded. He served as an Editor in Chief for European Alliance for Innovation (EAI) Transactions on Pervasive Health and Technology. He is serving as an associate editor for top ranked international journals such as IEEE Communications Magazine, IEEE Network, Future Generation Computer Systems, and IEEE Access. He served/serving as a guest editor for about two dozen special issues in journals such as IEEE Communications Magazine, IEEE Wireless Communications Magazine, Future Generation Computer Systems, IEEE Access, and Computer Networks. He has been involved in about one hundred peer-reviewed international conferences and workshops in various capacities such as a chair, co-chair and technical program committee member. He has been consecutively awarded with Outstanding Associate Editor of IEEE Access in 2018 and 2019 besides many others.

M. N. Marsono received the B.Eng. degree in computer engineering and the M.Eng. degree in electrical engineering from Universiti Teknologi Malaysia, in 1999 and 2001, respectively, and the Ph.D. degree in electrical and computer engineering from the University of Victoria, BC, Canada, in 2007. He is currently an Associate Professor in Electronic and Computer Engineering with the School of Electrical Engineering, Faculty of Engineering, Universiti Teknologi Malaysia. His research focuses on specialized hardware architecture and network algorithmics for high-throughput packet and flow processing. He works on dynamically reconfigurable platforms for middlebox, fog and edge computing, software-defined networking, and teletraffic engineering. He also works in domain-specific reconfigurable computing research, focusing on multicore/manycore system-on-chip, network-on-chip, design space exploration, mapping, and prototyping of the homogeneous and heterogeneous manycore SoC.