International Journal of Scientific Research in Computer Science, Engineering and Information Technology
ISSN : 2456-3307 (www.ijsrcseit.com)
doi : https://doi.org/10.32628/CSEIT22824

Comprehensive Study On EDGE-Cloud Collaborative Computing for Optimal Task Scheduling

[1]K.Vinothkumar, [2]Dr. D. Maruthanayagam
[1]Research Scholar, Sri Vijay Vidyalaya College of Arts & Science, Dharmapuri, Tamilnadu, India
[2] Head/Professor, PG and Research Department of Computer Science, Sri Vijay Vidyalaya College of Arts & Science, Dharmapuri, Tamilnadu, India

Article Info
Volume 8, Issue 2
Page Number : 75-90

Publication Issue :
March-April-2022

Article History
Accepted: 10 March 2022
Published: 20 March 2022

ABSTRACT

In recent years, Cloud and edge computing have got much attention because of the ever-increasing demands. There are many future technologies and advantages for systems to move towards clouds based on information keep methods. This includes a simple IT substructure and administration, and an effective distant approach from any place in the global with the steady computer network connections and efficient cost that cloud engineering can give. These paradigms impose to process the large amounts of generated data close to the data sources rather than in the cloud. One of the considerations of cloud edge based environment is resource management, which typically revolves around resource allocation, resource provisioning, task scheduling and improve performance. Aiming at the future problem of simulating service requests and optimal task scheduling during the operation of the cloud computing/edge computing environment, the real-time optimization scheduling technology of computing resources is studied, and elastic resource optimization scheduling is realized through data feature (quality) mining analysis, and collaborative resource management. Ensure that the simulation service quality meets the mission requirements and provide support. The main goal of this paper is to provide the better and deeper understanding regarding the scheduling approaches in the Edge-Cloud environment that covers the way in the scheduling approaches.

Keywords: Task Scheduling, Edge Cloud Computing, Optimization Algorithms, Cloud Data Centers and Resource Management.

I. INTRODUCTION

Cloud computing is regarded as a new paradigm of computing to improve network efficiency and satisfy the internet requirement with connected devices increasing by moving computation, control and data storage into the cloud [1]. However, cloud computing faces some challenges to some more stringent performance, such as latency and bandwidth, which are required by many application services. Edge

Copyright: © the author(s), publisher and licensee Technoscience Academy. This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License, which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited

computing as a new trend in computing is presented to increase infrastructure efficiency with low-latency and bandwidth-efficient services [2]. In computing environment, tens of billions of mobile devices are deployed on network edge and the computation is conducted near the end user (EU), whose processor speeds are increasing exponentially. The Cloud is a well tested and used solution that can extend the resource capabilities of the end devices with powerful data center topologies. Besides, Cloud is well equipped with the appropriate automation tools and platforms in order to offer the necessary transparency to the end devices, while hiding the complexity and the logistic details of this resource extension [3].

In a cloud computing paradigm, users can rely on extremely **rich storage and computing resources of a cloud computing center to expand the computing** and storage power of devices, and achieve the rapid processing of computing intensive tasks. Yet there are some disadvantages in the device cloud collaboration mode, such as incurring high transmission delay and pushing network bandwidth requirement to the limit. **In order to solve the problems of cloud computing for data processing, edge computing is put forward to provide desired computing** services for users by using computing, network, storage and other resources on edge, that is near a physical entity or data source. Edge computing moves the services and functions originally located in the cloud to the proximity of users, which integrates the cloud computing platform and the network to provide powerful computing, storage, networking, and communication capacity at the edge of the network. Compared with cloud computing, some applications of users in edge computing can be processed on an edge server near intelligent devices, thus significantly reducing data transmission delay and network bandwidth load required in edge-cloud collaboration.

By combining the benefits of cloud computing and edge computing, edge-cloud computing is one of the most promising ways to address all of the problems for improving the battery lifetime and application performance for user devices. Edge-cloud computing [4] performs each task on a user device, an edge or a cloud, which can provide better computing performance and transmission performance compared with edge computing or cloud computing in overall. While the **task offloading is one of the most challenge problems** must be addressed for improving the resource utilization efficiency in edge-clouds [5].

II. EDGE COMPUTING

Edge computing refers to the enabling technologies allowing computation to be performed at the edge of the network, on downstream data on behalf of cloud services and upstream data on behalf of IoT services. Here we define "edge" as any computing and network resources along the path between data sources and cloud data centers. ***For example, a smart phone is the edge between body things and cloud, a gateway in a smart home is the edge between home things and cloud,*** a micro data center and a cloudlet [6] is the edge between a mobile device and cloud. The rationale (validation) of edge computing is that computing should happen at the proximity of data sources. From our point of view, edge computing is interchangeable with fog computing [7], but **edge computing focus more toward the things side, while fog computing focus more on the infrastructure side.** We envision that edge computing could have as big an impact on our society as has the cloud computing. Figure 2 illustrates the two-way computing streams in edge computing. In the edge computing paradigm, the things not only are data consumers, but also play as data producers. At the edge, the things can not only request service and content from the cloud but also perform the computing tasks from the cloud. ***Edge can perform computing offloading, data storage, caching and processing, as well as distribute request*** and delivery service from cloud to user. With those jobs in the network, the edge itself needs to be well designed to meet the requirement efficiently in service such as reliability, security, and privacy protection [8].

Edge computing brings the resources of computation, storage, and networking significantly closer to the devices, users, and applications. It is viewed as one of the significant technologies associated with the Internet of Things and artificial intelligence in the next generation of networks [9]. It enables the data to travel at remarkably high speed without transferring data to the cloud or data centre. Edge computing is multilayered distributed architectures that balance the workload amidst edge layer, edge cloud, edge network and enterprise layer. **Edge computing comprises three nodes: Device edge, Local edge, and Cloud.** Figure 3 illustrates the overview of edge architecture with the relevant components within each node. Industry solutions/Apps can be considered as duplicate components that exist both in the device edge and local edge. Specific workloads are integrated either with the device edge and local edge, and they can be dynamically migrated amid of the nodes either manually or automatically. Local edge node manages multi-cloud management and network services.

- The device edge is the place where the device is occupied. These devices include cameras, sensors and any physical devices gathering data and interact with edge data. Some applications such as AI models with deep learning, video analytics are managed on these edge devices.
- Local edge encloses both applications and the network. The applications of the device edge reside in this node. The non-runnable applications on the device edge are executed at local edge, Such as IoT processing. Some virtualized and containerized network layers will run on this local edge.

Cloud is an environment where everything is brought together. It can either run at premises or as a public cloud. The applications that are not able to run in either of the nodes are handled by this cloud using orchestration layers Edge computing can also be defined in terms of computing power and latency. It allows us to perform necessary computations with low latency with high computing-intensive calculations. **Cloud, edge node, edge gateway and edge devices are the most essential layers in handling application** and transmitting data at the edge devices.

2.1. Architecture of Edge Computing

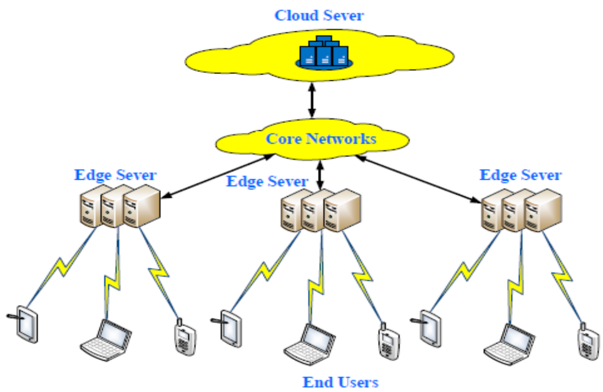

Figure-1: Edge Computing architecture

Edge computing, located between the mobile end devices and cloud data center, provides computing, storage, and network services. Edge computing has the ability of decentralized computation and storage compared with cloud computing. The resources near the user can support mobile devices for real-time communication. ***The main goal of edge computing is to preprocess data, reduce the delay, and serve the applications of low delay and real-time response.*** Currently, the common edge computing architecture is three-tier network architecture, as shown in Figure 1. As the underlying node, the end devices not only consume data but also produce data. In particular, mobile devices will require more resource from edge servers rather than cloud data center. The end devices include all IoT devices, such as smart mobile phone, intelligent vehicle, and virtual sensor nodes. These devices are able to sense data, communicate with the upper layer through sensor networks, 3G, Wi-Fi, and so on, and transmit the collected raw data. Compared with the servers in data centers, most of the IoT devices or sensors at the edge are somewhat limited in both computing power and battery capacity [10]. Edge devices include some traditional network devices (routers, switches, etc.) and some specially deployed devices (local servers). **Edge devices have the ability of computing, processing, storage, and forwarding data to the cloud servers.** Edge servers can be connected directly to nearby mobile devices via **one hop** wireless link. Micro Data Center **(MDC)** resources include computing resources, network resources, storage resources, and software resources. On the one hand, resources come from the cached local resources; on the other hand, they come from the data obtained by the monitoring equipment. As the top layer, cloud computing layer includes Data Center **(DC)** and some cluster servers. The cloud servers can receive data sent by edge devices for processing and storage [11].

The **three tier** architecture as follows:

End device: In different works, end-devices are also called MDs or mobile users (MUs). Various things can perceive and have certain storage and computing capability. Things continuously generate and collect multiple types of data can be processed locally, or be offloaded to the edge and the cloud. In the edge computing network, there are no only static end-devices (e.g., sensors in smart homes, video cameras in public places) but also dynamic ones such as UAVs and vehicles, making the resource management even more challenging.

Edge Server: ***The edge server, as the core of the three tier architecture, is an intermediate layer*** between the end device layer and the cloud server. From the perspective of hardware composition, the edge layer consists of various networking and computing equipment, such as cellular tower, edge server (ES), roadside unit (RSU), gateway, edge controller, etc. The edge layer provides wireless access to smart devices through the radio access technology, such as Long Term Evolution (LTE), Wireless Fidelity (Wi-Fi), and Dedicated Short-Range Communications (DSRC). Basically, the edge layer can provide more powerful storage and computing capabilities than the End device. From the perspective of software composition, the edge server has edge management capabilities that offer service orchestration and invocation and schedule the ESs to complete tasks. The edge layer can receive, process, and forward data streams from the end device layer, and achieve intelligent sensing, privacy protection, data analysis, intelligent computing, process optimization, and real-time control. Besides, since the edge and the cloud are complementary and mutually reinforcing, services in the cloud can be offloaded to the edge server for load balancing and better QoE. With the objective of reducing bandwidth usage and energy consumption of the cloud network (CN) as well as reducing the communication overhead between the edge and the cloud, the edge server is expected to schedule edge resources to enable rapid service response.

Cloud layer: The cloud layer consists of the existing cloud computing infrastructures, such as computing units, storage units, and micro data centers (MDCs), connected with the edge server through the CN (a.k.a, backbone network) Among the three layers, the cloud layer is undoubtedly the most powerful data processing and storage center. While ESs in the edge layer can process large amounts of data to reduce latency and energy consumption, the edge computing paradigm still requires the computing power and high-capacity storage infrastructure of the cloud to handle some tough tasks and global information. For example, the cloud layer can receive data streams from the edge layer, and send control information to the edge layer, and then from the edge layer to the End device, thereby optimizing the resource scheduling and field production process from a global perspective. Besides, based on the network resource distribution, the cloud layer can also dynamically adjust the deployment strategies and algorithms. Furthermore, it also provides decision-support systems, intelligent production, networking collaboration, service extension, personalized and customized service, and other domain-specific application services [12].

2.2. Edge Computing Applications

Medical Applications: "HPC Planning and Commissioning" HPC at the Edge for medical imaging merges HPC and medical sensing technology in order to provide precision medicine through the use of real-time advanced monitoring and analysis of a patient's medical data to detect early pathologies while lowering the risk of privacy breaches by keeping the data on site. This granular, yet massive amount of patient data can be analyzed at the Edge, transformed, and then only pertinent data is sent to the cloud such as alerts or data stripped of information that could lead to the patient's privacy being compromised. Medical Imaging at the Edge using HPC removes the latency and dependence on Cloud Computing resources, as well as reduces the patient's digital footprint by limiting how many systems have access to data.

Smart City

In the future cities will have sensors that will collect various data, for example, in transportation, medical health, and urban security. Moreover, urbanization is rapidly increasing. According to the UN, it is estimated that, by 2050, over 6 billion people will be living in the cities [13]. In the future, to have sustainable development in the town, a smart city is an excellent solution. This might help to solve the problems that may arise in food supply, medical care, transportation, culture and entertainment in the cities. **These sensors will usually generate a large volume of data, and this data should be processed quickly.** Sending these data to the cloud will need faster data movement (latency and data traffic in the network), and privacy. Therefore, these generated data should be processed closer to where it is produced. In general, ***Edge devices have limited computing and storage, so it is also necessary to integrate multiple computing model.***

Industrial/Manufacturing Applications

Industry 4.0 use at the Edge HPC in industrial automation environments. It aims towards waste reduction, work reduction, and worry reduction in the work space. It is used for connecting machines-to-machines and machines-to-people in a way where on-demand production environments, equipment, and workers can quickly and intelligently react to dynamically changing factory floor/environmental conditions [14]. Certain industrial applications may need to react quickly to real-time changing environmental conditions which may be uncovered in data too voluminous to be sent to the cloud, such as image recognition data that guides a robotic arm to interact with an object on a moving assembly line or creates alerts if dangerous conditions arise. Moving data offsite for analysis may also incur transmission latencies that exceed the reactions times required for industrial applications, such as being able to shutdown an assembly line if a foreign object interferes with the industrial process. This is all assuming that the industrial site is even able to acquire a high-speed network connection due to geographic constraints [15].

Smart Transportation

Autonomous vehicles are embedded with a large number of sensors such as cameras, lidars, etc. Uploading sensor data to the cloud computing center will increase the difficulty of real-time processing. Therefore, performing edge computing on the car can speed up processing and enhance the real-time feature of road environment decisions. The power supply of the drone itself is limited. If the data is transmitted to the cloud center, the energy consumption is high, and the delay is high. ***Drones generate a large amount of high-definition video data, making it difficult to achieve real-time transmission of wireless networks*** and commands from receiving centers [16]. Edge computing solves these problems well. The edge end processes the data sensed by the drone, reducing the energy consumption of data transmission and ensuring real-time performance.

Smart Home

With the popularization of the Internet of Everything application, home life has become more and more intelligent and convenient. However, the practice of connecting smart lighting systems, smart TVs, smart robots and other devices to the cloud computing center through the Wi-Fi module is far from complete. Meet the needs of smart homes. In the smart home environment, a large number of wireless sensors and controllers are deployed in the rooms, pipes, floors and walls. For the consideration of data transmission load and data privacy, the processing of these sensitive data should be completed within the home. Traditional cloud computing models it is no longer fully applicable to smart home applications, and the edge computing model makes it easier to connect and manage smart home devices within the home, and processes the data generated by these devices locally, reducing the load of data transmission bandwidth [17].

VR / AR

Both **VR**(Virtual Reality) and **AR**(Augmented Reality) **technologies need to collect real-time information related to user status including user position and orientation,** then perform calculations and process them based on the calculation results [18]. The server can provide it with rich computing and storage resources, cache the audio and video content that needs to be pushed, and based on the one-to-one correspondence between positioning technology and geographic location information, determine the pushed content based on the location information, and send it to the user or quickly simulate a three-dimensional dynamic Visualize and interact with users [19].

2.3. Edge computing Key Features

- Very low latency.
- Reduced bandwidth limit.
- Flexibility in deployment.
- Automation.

Edge computing is the main domain which covers a family of related technologies as sub domains such as cloudlet, mobile cloud computing, and multi-access edge computing. All these sub-domains are based on the concept that computational resources should be Present at the edge where network encounters the real demand challenge, which leads to:

1. Gain of fast access to the computation resources by the end users.
2. Reducing the traffic load on network core significantly [20].

Although cloud computing platforms can be used to store data and compute, centralized cloud computing has exposed its problems in the face of massive user device connections, explosive growth in data traffic and users increasing demand for better service quality. First, the real-time is not enough. In addition, the transmission bandwidth between the cloud and edge devices is insufficient [21].

2.4. Technology Move from Cloud to Edge Computing

Edge defines as any computing and network resource along the path between data sources and data centers in the cloud. **A Smartphone, for example, is the edge between a human body and a cloud, a smart home gateway is the edge between a home and a cloud,** and a micro data center is the edge between a mobile device and a cloud. The reason for the development of this type of advanced computing is that computing should ideally occur in close proximity to the data source. From our point of view, the term edge computing is interchangeable with fog computing, but edge computing focuses more on the things side, whereas fog computing is concentrated more on the infrastructure side. We envision that edge computing could have as big an impact on our society as cloud computing. ***The utilization of edge computing means that part of the data processing function can be carried out closer to the source of the data;*** rather than first sending it over the internet to a datacenter for processing in the cloud. The edge computing approach can generate several benefits for an organization. **First, it can maximize big data analysis by keeping some tasks out of the main cloud's data storage queue, allowing them to complete faster.** It is simply not necessary for each analytical action to pass from the device to the cloud and vice versa. In the case of a solution with video cameras for example for retail, security applications or consumer behavior, it is possible to create business rules that only send reports to the data center about certain events because edge computing devices can sort tasks according to priority, thereby keeping critical actions inside the node. For instance, in the retail case, it is possible to make video surveillance cameras start to transmit data to the cloud only when a defined type of event occurs; for instance, when an intruder or shoplifter has been identified, transmission is initiated [22].

The following requirements propel computing to move to the edge and they explain why edge computing is an essential model [23, 24].

Push from cloud services: Placing all computing tasks on the cloud has proved to be an effective method for data processing, because the computing power on the cloud surpasses the capacity of the things on the edge. However, the network suffers a deadlock compared to the speed of the rapidly developing data processing. As the amount of data generated at the edge continues to increase, the speed of data transportation impedes the traffic flow in the cloud-based computing model. **Two examples can be given here:** The first example, states that a **Boeing 787 generates about 5 gigabytes of data every second,** but the bandwidth between the aircraft and either the satellite or the base station on the ground is insufficient for data transmission. **The second example in states that a vehicle generates 1 gigabyte of data every second,** and that the vehicle needs real-time processing in order to make correct decisions. Sending all data to the cloud for processing would ineffectively prolong the response time. It would also pose a challenge on the reliability of the current network bandwidth and its ability to support a large number of vehicles in one area. Thus, processing data at the edge ensures less response time and network pressure as well as better processing. Pull from Internet of things (IoT): In the near future, most electric devices will be included in the IoT, and they will play the roles of both data producers and consumers alike, such as air quality sensors, Light-Emitting Diode **(LED)** bars, streetlights, and even an Internet-connected microwave oven we can conclude that, in a few years, the number of things on the edge will exceed billions. Thus, they will produce massive raw data that the standard cloud computing model will not be able to handle. **This leads to the fact that most IoT data will not make its way to the cloud, and will rather be consumed at the network edge.** The classical structure of the cloud computing model is inadequate for IoT for many reasons: First, the amount of data at the edge is too large and it utilizes a lot of unnecessary bandwidth and computing resources. Second, the requirement for privacy protection hinders cloud computing in IoT. Third, energy constraints restrict most of the IoT end nodes, and the wireless communication module usually consumes a lot of energy, so it might be more energy efficient to discharge some computing tasks to the edge.

Change from data consumer to producer: The front-end devices at the edge usually act as a data consumer in the cloud computing model, such as watching a YouTube video on your smart phone, but nowadays people are also generating data via their mobile devices. Switching from data consumer to data producer/consumer requires more peripheral placement. For instance, people ***nowadays normally take pictures or record videos and then share the data via a cloud service such as Twitter, Snapchat, and Instagram.*** Twitter users sent around 473,400 tweets per minute in 2018; Snapchat users shared 2,038,333 snaps; and Instagram users posted 49,380 new photos in that year . **However, uploading a large image or a video clip would occupy a lot of bandwidth,** and this requires adjusting the video clip to an appropriate resolution before uploading it on the cloud. Wearable health devices can also be an example. As physical data collected by the things at the edge of the network are often privately owned, **user's privacy could be better protected by processing data at the edge rather than uploading them** on the cloud.

Decentralized cloud and low-latency computing: Centralized cloud computing is not always the ideal strategy for geographically distributed applications. To enhance the service provided, the computing must be done closer to the data source. For any web-based application, this benefit can be generalized. Recently, users have become more interested in using location-aware applications, and this entices (attract) application providers to improve their service by computing on edge nodes that are closer to users. Edge devices generate many data streams, and performing the analytics on a distant cloud obstructs real-time decision-making. Using current cloud infrastructures in real-time applications causes serious latency issues between an edge device and the cloud, and any potential delay would not be in line with the requirements of latency-sensitive applications. Video streaming generates most of the mobile traffic, and it is still difficult to maintain user's satisfaction when the majority of the network traffic flows to the same source. In the same way, gamers face similar latency obstacles when using multimedia applications [20]. **In this case, employing edge nodes closer to users can minimize network latency** to enhance the computations performed on the cloud.

Surmounting resource limitations of front-end devices: User devices have somewhat restrained hardware resources when compared to compute and/or storage resources of the servers hosted within a cloud data centre. Front-end devices can be mainly divided into two types: devices carried by people and others placed in the environment. The main role of the two front-end device types is to obtain real-time data by capturing sensory input in the form of text, audio, video, touch, or motion, and then cloud services process the transferred data. These devices, however, are unable to perform complex analytics due to their hardware limitations. Therefore, data often needs to be transmitted to the cloud to meet computational processing requirements and to send back useful information to the front-end device. However, the service does not have to use all the data it receives from the front-end device in order to build analytical cloud workloads. ***Data can possibly be filtered or even analyzed at edge nodes, which may have spare computer resources for managing data.***

Sustainable energy consumption: Numerous energy-consuming data centers have been recently established worldwide. However, the financial and environmental impact of the high energy consumption at these data centers is a major issue that needs to be tackled, and enhancing the energy efficiency of such data centers is a main challenge in the cloud computing model. As the number of applications moving on to the cloud is increasing, the growing energy demand may become unsustainable [25].

2.5. Characteristics of Edge computing

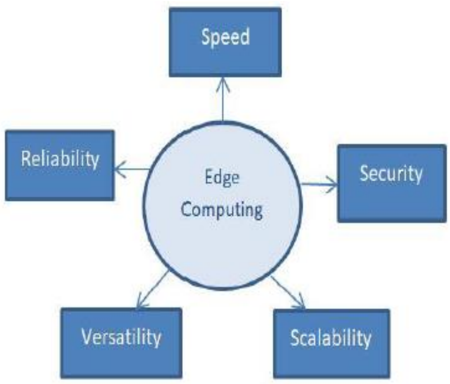

Figure 1 (a): Characteristics of Edge Computing

Speed: There are many companies that consider speed as very important factor for their core business. For example, **in healthcare system, losing even a fraction**

of second could be life critical. In data-driven businesses lagging speeds sometimes frustrate customer. ***Edge computing helps to reduce latency and enhances the performance of network by processing data*** locally on edge devices.

Security: Traditional cloud computing technology is more prone to denial of service attack due to its centralized architecture. Other side, edge computing distributes application, storage and processing across different range of devices. Since mostly data rather than transmitting back to central data server will be processed on local servers and eventually reduces the risk of compromising the data.

Scalability: As the companies grow, it is no longer required to establish private and centralized data centers. Due to **edge computing, it is much easy for the companies to scale their operations.** The idea of not relying centralized infrastructure that allows them to scale their data and computing quickly and more efficiently.

Versatility: Different companies can effortlessly target preferable markets without investing much in expensive infrastructure. **Edge data centers could efficiently provide services to end users with low latency.** Edge computing helps IoT devices to accumulate large amount of actionable data.

Reliability: In edge computing, **edge data centers are positioned much closer to end users that inhibit less chances of network problem.** There are multiple edge nodes that are connected to network and if any failure occurs that would not let interrupt the whole services. There are multiple ways through which that can be rerouted. Edge architecture that comprises of edge computing devices and data centers can help to provide unparalleled reliability [26].

III. EDGE-CLOUD COLLABORATIVE COMPUTING MODEL

In edge-cloud collaborative computing, **the edge is responsible for data computing and storage in the local area, and the cloud is responsible for big data analysis, mining, and algorithm training optimization.** The edge-cloud collaboration computing model is shown in Figure 2. The collaboration of the edge-cloud can be divided into two parts. ***The first is functional collaboration.*** This kind of collaboration assumes different functions based on different geographic spaces and roles of different computing devices. For example, the edge is responsible for preprocessing, and the cloud is responsible for multi-channel data processing and service provision.

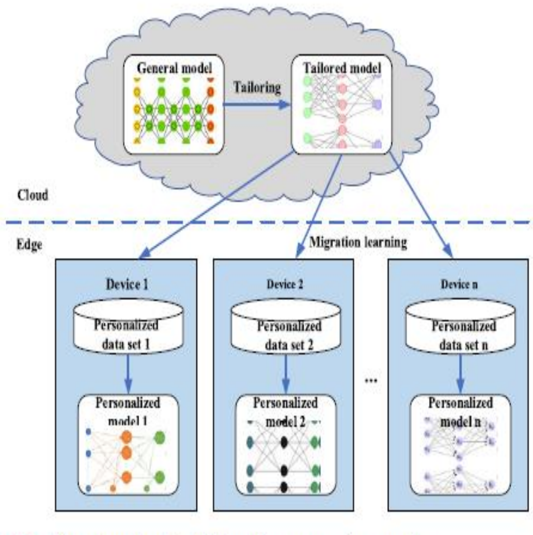

Figure 2 : The edge-cloud collaborative computing mode.

The second is performance collaboration. This is due to the limitation of computing power, and computing devices of different levels undertake tasks with different computing power requirements, including longitudinal cutting and assignment of tasks [27].

Resource allocation refers to the efficient allocation of constrained resources to competing services with various features and requirements, in a way that edge system obtains maximum resource utilization and also satisfies the services. ***There is a verity of resource***

allocation models proposed for edge computing which aim to optimize resource utilization, profit, power efficiency, QoS (e.g. response time) [28].

In adaptive of a task scheduling mechanism for high concurrency intelligent applications in a cloud-edge collaborative environment. The network topology diagram of the cloud-edge system this study work is composed of the user terminal devices (e.g., computers, smart phones, smart watch), edge nodes, relay management nodes, and cloud data center, as shown in Figure-3. Among them, the relay management node [29] is more difficult to understand, which has **two functions in the cloud-edge system. One is to collect information about the resources** and status of each node, including node storage, computing resource information, tasks being executed, and tasks waiting to be executed. **The second function is task collection and scheduling.**

The edge node forwards the request to the relay management node when it receives the task request. The relay management node allocates tasks requests to the corresponding node to process according to the collected information and the built-in task scheduling algorithm. Based on the constructed network topology diagram, this study concludes the scheduling strategy aiming at intelligent applications requests and deployed them in the relay management node.

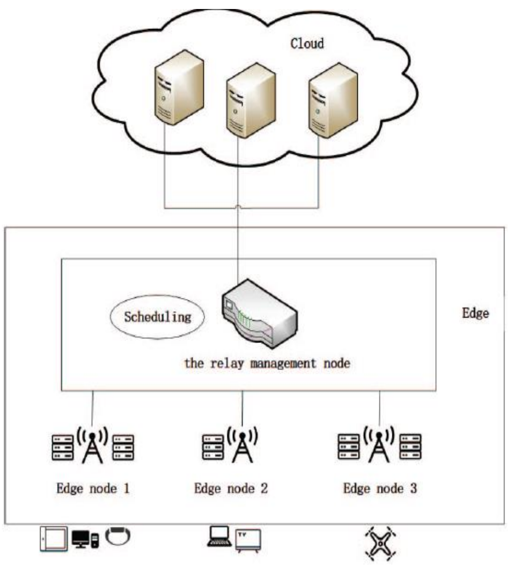

Figure 3 : The cloud-edge collaborative system model.

Tasks can be scheduled to the edge or the far cloud based on energy consumption and time delay. **For the problem that needs to be processed in the cloud center, how to perform** proper scheduling to achieve the goal is worthwhile research question. In the scheduling process, various performance-based performance indicators such as system utilization, execution time, load balance, network communication cost, delay and that like are used [30]. The heuristic task ***scheduling algorithm can easily schedule tasks and provide the best solution.*** However, it does not guarantee the best results and is easy to fall into partial selection. The metaheuristic algorithm is an improved algorithm based on a heuristic algorithm, which is a combination of random algorithms and local search algorithm [31]. It enables the exploration and development of search space and handles a large amount of search space information.

In addition, it can use learning strategies to acquire and master information to effectively find approximate optimal solutions. Among them, genetic algorithm (GA), particle swarm optimization (PSO), and ant colony algorithm (ACO) are the most widely used evolutionary algorithms in the task scheduling in recent years. **However, these algorithms usually converge prematurely and are prone to finite optimally.** When approaching the optimal solution, it may also swing left and right, making the convergence slower.

The traditional scheduling strategies of edge computing tasks are to offload all computing-intensive tasks of edge devices to an edge server for processing [32]**. However, it may result in the waste of computing and storage sources in edge devices and cloud computing centers**. In addition, many devices may access an edge server at the same time period. As a result, the server may face too many computing tasks, thus resulting in a long queue of tasks. This increases the completion time of all queued tasks, even causing the processing delay of tasks in the edge server to exceed that at the edge devices. On the other hand, many edge devices may be idle, resulting in a waste of their computing resources; and resource-rich cloud centers may be underutilized.

To solve the above problems, we can combine a cloud center, edge servers and edge devices together to efficiently handle the computing tasks of edge devices **via task offloading.** According to the computing tasks characteristics, optimization objectives and system status, we should utilize the computing and storage resources of a cloud center, edge servers and edge devices, and schedule computing tasks to them for processing on demand. It can effectively reduce the load of edge servers and improve the utilization of resources, and reduce the average completion time of computing tasks in a system [33].

3.1. Resource Management Techniques in Cloud Edge Scenarios

Resource Allocation: Resource allocation represents a technique that is used **to optimize the utilization of resources and reduce the required costs** for processing [34]. Fulfillment time of a task is an important aspect that should be considered since it can impact the completion of resource allocation [35].

Workload Balance: Workload balancing is an important factor used **to manage energy effectiveness and also avoid congestion, low-load resource management, and overload.** Currently, this represents a challenge for the processing nodes, which are placed in the fog environment. For instance, in [36], a workload balancing algorithm is proposed for fog computing, aiming to reduce the data flow latency in the transmission procedures by connecting IoT devices to the appropriate base stations (BSs).

Resource Provisioning: Resource provisioning is a further step in resource allocation. As discussed above, ***resource allocation deals with just assigning a set of resources to a task,*** while resource provisioning deals with the activation of the allocated resources.

Task Scheduling: To manage a large set of tasks that are working together and are dependent on a certain set of resources, task scheduling algorithms have been proposed to define a schedule to service tasks **to avoid conditions such as deadlocks** [37].

3.2. Resource scheduling in Edge Computing

Server placement: The location of edge servers could significantly influence on the efficient deployment of edge environments such as mobile edge computing and smart cities. A proper placement of edge servers across large-scale networks leads ***to reducing access delays and improving resource utilization of edge servers.*** In server placement, the challenge is how to select the best location for each edge server, and how to assign base stations to edge servers.

Service caching: It refers to maintaining various sets of cloud services on the edge servers to **accelerate short-distance connections and faster access to the services** requested by users/devices. Since the efficiency of the resource allocation is affected by the set of cached services, the challenge is how to optimally select them among all cloud services for each edge server in order to improve user-experience and regarding resource constraints of the servers.

Service placement: In most cases, edge servers cooperate together to provide required services for different applications. Using service chain architecture, service providers can decouple services into several micro services in a specific order each of which responsible for a part of a given service. ***These micro services can be placed on edge services while they need to communicate to each other.*** For instance, in Network Function Virtualization Infrastructures (NFVI) service function chains comprised of Virtualized Network Functions (VNFs) can be deployed on different edge/cloud nodes. Due to resource constraints of network links and servers, application heterogeneity, varying of demanded service performance and also the budget limitation to rent edge nodes, an optimal service placement is critical to satisfy both users and service providers.

Service Migration: In **MEC** due to dynamically moving the users (e.g. connected vehicles and smart phones), seamless service ***migration has become a crucial issue to ensure service continuity and better user experience.*** The conflict between mobility of users and limited coverage of edge servers may lead to performance degradation and even service interruption. The main challenge of service migration is to realize when and where an edge service allocated to a mobile user should migrate from current edge server to another

Service provider: In addition to users, the edge computing ecosystem incorporates multiple actors, such as edge infrastructure SPs, edge computing service providers, application service providers, and mobile network operators. Although these SPs and operators are resource-rich and have powerful serviceability, they are ***all commercial entities aiming at earning revenue by providing services.*** In this context, designing an appropriate resource scheduling strategy can help them get maximal revenue during service providing competition at a minimal cost.

Edge network: Edge resources are distributed and scattered in the edge network. It is a waste of resources if scattered ones cannot be efficiently utilized by resource scheduling. For example, the parked vehicles (PVs) account for a large portion of the global vehicles and have idle time to perform computational workloads [38]. If an efficient resource strategy is applied, they can be combined **to establish an available and cost-effective computing** resource pool, which helps to alleviate workloads of edge computing servers and promote the distributed computing environment.

3.3. Challenges & Issues and Research Techniques Edge Cloud

Discovering Edge Nodes: Discovering resources and services in a distributed computing environment is an area that is well explored. This is facilitated in both tightly and loosely coupled environments through a variety of techniques that are incorporated into monitoring tools [39] and service brokerages [40]. Techniques such as benchmarking underpin decision-making for mapping tasks onto the most suitable resources for improving performance. However, exploiting the edge of the network requires discovery mechanisms to find appropriate nodes that can be leveraged in a decentralized cloud set up. These mechanisms cannot be simply manual due to the sheer volume of devices that will be available at this layer. Moreover, they will need to cater for heterogeneous devices from multiple generations as well as modern workloads, for example large scale machine learning tasks, which were previously not considered.

Benchmarking methods will need to be significantly rapid in making known the availability and capability of resources. ***These mechanisms must allow for seamless integration (and removal) of nodes in the computational workflow*** at different hierarchical levels without increasing latencies or compromising the user experience. Reliably and proactively dealing with faults on the node and autonomically recovering from them will be desirable. Existing methods used in the cloud will not be practical in this context for the discovery of edge nodes [41].

i) SLA-aware server placement and service migration: Most focus in the previous studies on the resource allocation problem in edge computing has been on task offloading, service caching and service placement, and a few works have been done on server placement and service migration. However, optimally locating of a wide range of heterogeneous edge servers across the edge network with different domains and features becomes a big challenge of resource allocation in edge computing. In addition, ***service migration strategies are needed to cope with the high mobility of dynamic edge environments*** in order to satisfy the user expectations and also guarantee the required SLA. For this end, more studies can be conducted in this area.

ii) Resilient (flexible) resource allocation: Reliability and availability are two critical challenges that edge computing faces with. The movement ability of users, limited resources of nodes and the need for cooperation between cloud data centers and edge servers may cause a high risk of fault and failure in the edge infrastructure. Therefore, design and implementing a failure-aware and multi-holistic provisioning and allocation mechanism is an open issue which is neglected in the literature yet.

iii) Edge federation framework: There is still a lack of comprehensive federation framework (similar to hybrid cloud federation) **to monitor and control the resources across the edge network** by which resource sharing and allocation can be performed more effective.

iv) Large scale experiments: To gain more benefits from task offloading and resource allocation in edge networks, there is a need to conduct a variety of large scale experiments on the real edge services along with cooperation between edge servers.

v) Dynamic bandwidth allocation: Static bandwidth allocation is mostly considered in the previous works, which leads to unfairness in resource sharing and allocation. Therefore, ***proposing dynamic bandwidth allocation approaches provide more effective techniques*** for the edge environments.

vi) Optimization criteria: To find an optimal solution for resource allocation problem, optimization metrics may influence on each other. For example, increasing utilization of resources in order to reduce energy consumption may lead to increasing latency, failure and unavailability of the edge networks. Because of that, optimization criteria should be prioritized based on different workloads and required QoS [42].

IV. CONCLUSION

In this paper, we have **reviewed the resource management and scheduling approaches in the Edge Cloud domain.** Edge or cloud location to improve the multiple key performance parameters. The main objective behind this paper is to **give a broader and deeper understanding regarding the scheduling approaches in the Edge Cloud environment** that paves the way in the scheduling approaches. The core reason behind the scheduling algorithm is to schedule a workload/ task at the intended edge or cloud location to improve the multiple key performance parameters. First, overview by elaborating on what exactly resource scheduling refers to in edge computing. Second, researcher presents the architecture and different collaborative manners for edge cloud applications. Third, researcher discussed the research issues and outline of task scheduling techniques is presented, which is the prominent (important) effort of this projected research work.

V. REFERENCES

[1]. Freeman H, Zhang T, "The emerging era of fog computing and networking" [The President's Page]. IEEE Communications Magazine, 2016, 54(6): 4-5.

[2]. Chang H, Hari A, Mukherjee S, et al. Bringing the cloud to the edge. Computer Communications Workshops. IEEE, 2014: 346-351.

[3]. Firdose Saeik, Marios Avgeris, Dimitrios Spatharakis, Nina Santi, Dimitrios Dechouniotis,"Task Offloading in Edge and Cloud Computing: A Survey on Mathematical, Artificial Intelligence and Control Theory Solutions", 2021.

[4]. Z. Xu, W. Liang, M. Jia, M. Huang, and G. Mao, ``Task offloading with network function requirements in a Mobile Edge-Cloud Network,'' IEEE Trans. Mobile Comput., vol. 18, no. 11, pp. 2672_2685, Nov. 2019.

[5]. J. Ren, D. Zhang, S. He, Y. Zhang, and T. Li, ``A survey on end-edge cloud orchestrated network computing paradigms: Transparent computing, mobile edge computing, fog computing, and cloudlet,'' ACM Comput. Surv., vol. 52, no. 6, Oct. 2019.

[6]. M. Satyanarayanan, P. Bahl, R. Caceres, and N. Davies, "The case for VM-based cloudlets in

mobile computing," IEEE Pervasive Comput., vol. 8, no. 4, pp. 14–23, Oct./Dec. 2009.

[7]. Open Fog Architecture Overview. Open Fog Consortium Architecture Working Group. Accessed on Dec. 7, 2016. [Online]. Available:http://www.openfogconsortium.org/wp-content/ uploads/OpenFog-Architecture-Overview-WP-2-2016.pdf.

[8]. Weisong Shi, Fellow, IEEE, Jie Cao, Student Member, IEEE, Quan Zhang, Student Member, IEEE, Youhuizi Li, and Lanyu Xu"Edge Computing: Vision and Challenges", IEEE Internet of Things Journal, vol. 3, no. 5, October 2016.

[9]. Jararweh, Y., Doulat, A., AlQudah, O., Ahmed, E., Al-Ayyoub, M. and Benkhelifa, E., 2016, May. The future of mobile cloud computing: integrating cloudlets and mobile edge computing. In IEEE-2016 23rd International Conference on Telecommunications (ICT) (pp. 1-5).

[10]. J. Pan and J. McElhannon, "Future edge cloud and edge computing for internet of things applications," IEEE Internet of Things Journal, 2017.

[11]. Guangshun Li , Jianrong Song , JunhuaWu , and Jiping Wang,"Method of Resource Estimation Based on QoS in Edge Computing", Hindawi Wireless Communications and Mobile Computing Volume 2018, Article ID 7308913, 9 pages.

[12]. Quyuan Luo, Shihong Hu, Changle Li, Senior Member, IEEE, Guanghui Li, and Weisong Shi, Fellow, IEEE "Resource Scheduling in Edge Computing: A Survey".

[13]. Wuhui Chen, Zhen Zhang, and Baichuan Liu. Smart cities enabled by edge computing. In Edge Computing: Models, technologies and applications, pages 315-337, 2020.

[14]. Jay Lee, Hossein Davari, Jaskaran Singh, and Vibhor Pandhare. Industrial artificial intelligence for industry 4.0-based manufacturing systems. Manufacturing letters, pages:20-23, 2018.

[15]. Ezhilmathi Krishnasamy, Sebastien Varrette, Michael Mucciardi,"Edge Computing: An Overview of Framework and Applications", Journal of prace, 2018.

[16]. Weisong Shi, Hui Sun, Jie Cao, Quan Zhang and Wei Liu 2017 Edge Computing New Computing Model in the Era of IOT Computer Research and Development 54(5) pp: 907-924.

[17]. Zishu Li, Renchao Xie, Li Sun and Tao Huang 2018 Mobile Edge Computing Survey Telecommunications Science pp: 87-101.

[18]. Fengbin Zheng, Dongwei Zhu, Wenqian Zang, Jinlin Yang and Guanghui Zhu 2020 Edge Computing-A New Computing Paradigm Review and Application Research Frontiers of Computer Science and Technology pp: 208-225.

[19]. Wei Cui, "Research and Application of Edge Computing Based on Deep Learning", Journal of Physics: Conference Series, 1646 (2020) 012016 IOP Publishing doi:10.1088/1742-6596/1646/1/012016.

[20]. Ahmed M. Alwakeel,"An Overview of Fog Computing and Edge Computing Security and Privacy Issues", Sensors 2021, 21, 8226. https://doi.org/10.3390/s21248226.

[21]. Shida Lu1, Rongbin Gu1, Hui Jin2, Liang Wang1, Xin Li 2, (Member, IEEE), and Jing ," Qos-Aware Task Scheduling In Cloud-Edge Environment",Ieee Access,Vol 9,2021.

[22]. Dr. Khalaf Khatatneh1 Osama Nawafleh2 Dr.Ghassan Al-Utaibi, "The Emergence of Edge Computing Technology over Cloud Computing", International Journal of P2P Network Trends and Technology (IJPTT) - Volume 10 Issue 2 – Mar – April 2020.

[23]. ShiW, Cao J, Zhang Q,Youhuizi L, and Lanyu X. Edge computing: vision and challenges. IEEE Internet of Things Journal. 2016:3(5):637–646.

[24]. Varghese B, Wang N, Barbhuiya S, Kilpatrick P, and Nikolopoulos D. Challenges and

opportunities in edge computing. In: 2016 IEEE International Conference on Smart Cloud (SmartCloud); 2016. pp. 20–26.

[25]. Auday Al-Dulaimy1, Yogesh Sharma1,Michel Gokan Khan1 and Javid Taheri,"Introduction to edge computing",0 DOI:10.1049/PBPC033E_ch1,2019.

[26]. Jyotsna1, Parma Nand2 , " Fog computing and Edge computing: An edge over cloud computing" Turkish Journal of Computer and Mathematics Education,Ï Vol.12 No.11 (2021), 4887-4894.

[27]. Chun Jiang And Jiafu Wan , "A Thing-Edge-Cloud Collaborative Computing Decision-Making Method For Personalized Customization Production", IEEE Access, Received December 14, 2020, Accepted January 5, 2021, Date Of Publication January 8, 2021, Date Of Current Version January 20, 2021.

[28]. Mohammad Ali Khoshkholgh, Michel Gokan Khan, Yogesh Sharma, Javid TaheriResource Allocation Models in/for Edge Computing, Chapter · September 2020, DOI: 10.1049/PBPC033E_ch7.

[29]. J. Fan, R. Li, and S. Li, "Research on task scheduling strategy: Based on smart contract in vehicular cloud computing environment," in 2018 1st IEEE International Conference on Hot Information-Centric Networking (HotICN), 2018, pp. 248–249.

[30]. T. Jena and J. R. Mohanty, "Disaster recovery services in inter cloud using genetic algorithm load balancer," International Journal of Electrical and Computer Engineering, vol. 6, no. 4, p. 1828, 2016.

[31]. Y. Xu, J. K. Li, and K. Li, "A genetic algorithm for task scheduling on heterogeneous computing systems using multiple priority queues," Information Sciences, vol. 270, pp. 255–287, 2014.

[32]. Zhao, J.; Li, Q.; Gong, Y.; Zhang, K., "Computation Offloading and Resource Allocation for Cloud Assisted Mobile Edge Computing in Vehicular Networks". IEEE Trans. Veh. Technol. 2019, 68, 7944–7956.

[33]. Shichao Chen, Qijie Li , Mengchu Zhou, Abdullah Abusorrah , "Recent Advances in Collaborative Scheduling of Computing Tasks in an Edge Computing Paradigm", Sensors 2021, 21, 779.

[34]. Rafique, H.; Shah, M.A.; Islam, S.U.; Maqsood, T.; Khan, S.; Maple, C. "A Novel Bio-Inspired Hybrid Algorithm (NBIHA) For Efficient Resource Management in Fog Computing". IEEE Access 2019, 7, 115760–115773. [Crossref].

[35]. Ni, L.; Zhang, J.; Jiang, C.; Yan, C.; Yu, K., "Resource Allocation Strategy In Fog Computing Based On Priced Timed Petri Nets". IEEE Internet Things J. 2017, 4, 1216–1228. [Crossref]

[36]. Fan, Q.; Ansari, N., "Towards Workload Balancing in Fog Computing Empowered Iot". IEEE Trans. Netw. Sci. Eng. 2020, 7, 253–262. [Crossref].

[37]. Adriana Mijuskovic 1, Alessandro Chiumento 1, Rob Bemthuis 1 , Adina Aldea 2 And Paul Havinga "Resource Management Techniques For Cloud/Fog And Edge Computing: An Evaluation Framework And Classification", Sensors 2021,21,1832. Https://Doi.Org/10.3390/S21051832.

[38]. X. Huang, R. Yu, J. Liu, and L. Shu, "Parked vehicle edge computing: Exploiting opportunistic resources for distributed mobile applications," IEEE Access, vol. 6, pp. 66 649–66 663, 2018.

[39]. J. Povedano-Molina, J. M. Lopez-Vega, J. M. Lopez-Soler, A. Corradi, and L. Foschini, "DARGOS: A Highly Adaptable and Scalable Monitoring Architecture for Multi-tenant Clouds," Future Generation Computer Systems, vol. 29, no. 8, pp. 2041–2056, 2013.

[40]. N. Grozev and R. Buyya, "Inter-Cloud Architectures and Application Brokering:

Taxonomy and Survey," Software: Practice and Experience, vol. 44, no. 3, pp. 369–390, 2014.
[41]. Blesson Varghese, Nan Wang, Sakil Barbhuiya, Peter Kilpatrick and Dimitrios S. NikolopoulosĬ, "Challenges and Opportunities in Edge Computing", 2016 IEEE International Conference on Smart Cloud.
[42]. Mohammad Ali, Michel Gokan Khan, Yogesh Sharma, Javid Taheri Khoshkholgh, "Resource allocation models in/for edge computing",September 2020, DOI: 10.1049/PBPC033E_ch7.

ABOUT THE AUTHORS

K.Vinothkumar received his **M.Phil** Degree from Periyar University, Salem in the year 2015. He has received his **M.Sc.,** Degree from Hindusthan College of Arts and Science, Coimbatore affiliated to Bharathiar University, Coimbatore in 2013. He is pursuing his **Ph.D** Degree (Part-Time) in Sri Vijay Vidyalaya College of Arts & Science, Dharmapuri, Tamilnadu, India. He is working as **HOD Cum Assistant Professor** in Department of Computer Science at Kavitha's College of Arts and Science (Co-Ed.,), Vaiyappamalai, Tiruchengode, Namakkal. His current research of interests includes Edge Computing, Cloud Computing, Mobile Computing and Ad hoc Networks.

Dr.D.Maruthanayagam received his **Ph.D** Degree from Manonmaniam Sundaranar University, Tirunelveli in the year 2014. He received his **M.Phil** Degree from Bharathidasan University, Trichy in the year 2005. He received his **M.C.A** Degree from Madras University, Chennai in the year 2000. He is working as **HOD Cum Professor**, PG and Research Department of Computer Science, Sri Vijay Vidyalaya College of Arts & Science, Dharmapuri, Tamilnadu, India. He has above **21 years** of experience in academic field. He has published **7 books**, more than **55 papers** in International Journals and **30 papers** in National & International Conferences so far. His areas of interest include Computer Networks, Grid Computing, Cloud Computing and Mobile Computing.

Cite this article as :

K. Vinothkumar, Dr. D. Maruthanayagam, "Comprehensive Study On EDGE-Cloud Collaborative Computing for Optimal Task Scheduling", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 8 Issue 2, pp. 75-90, March-April 2022. Available at
doi : https://doi.org/10.32628/CSEIT22824
Journal URL : https://ijsrcseit.com/CSEIT22824