International Journal of Information Technology and Web Engineering
Volume 17 • Issue 1

On Cost-Aware Heterogeneous Cloudlet Deployment for Mobile Edge Computing

Hengzhou Ye, Guilin University of Technology, China
Fengyi Huang, Guilin University of Technology, China*
Wei Hao, Guilin University of Technology, China

ABSTRACT

Edge computing undertakes downlink cloud services and uplink terminal computing tasks; data interaction latency and network transmission cost are thus significantly reduced. Although a lot of research has been conducted in mobile edge computing (MEC), which assumed that all homogeneous cloudlets are placed in WMAN and user mobility is also ignored, little attention has been paid to how to place heterogeneous cloudlets in wireless metropolitan area network (WMAN) to minimize the deployment cost of cloudlets. Meanwhile, the method of selecting an optimal access point (AP) for deployment, modeling, and heuristic algorithm (HA) needs to be improved. Therefore, this paper designs a new heterogeneous cloudlet deployment model considering the quality of service (QoS) and mobility of users, and the improved heuristic algorithm (IHA) is proposed to minimize cloudlet deployment cost. The extensive simulations demonstrate that IHA is more efficient than HA, and the designed model is superior to the existing work.

KEYWORDS

Cloudlet Deployment Cost, Heterogeneous, Latency, Minimization, Mobile Edge Computing (MEC), Optimal AP, QoS, Users Mobility

INTRODUCTION

Benefited from the rapid development of wireless network technology, smart mobile devices, mobile device software and hardware technologies, the growing number of users are peculiarly prone to run related services on mobile devices than on traditional computers. However, portable smart mobile devices are limited by enhanced computing resources, including computing ability, communication resources, storage and usability functions, including power, size, and weight. Meanwhile, it is difficult to provide computing resource demands for intensive and complex user tasks. Therefore, there is an increasing need for mobile users offloading tasks to the cloud, which has given birth to the new paradigm of Mobile Cloud Computing (MCC) (Gai et al., 2016; Pang et al., 2017; Shaukat et al., 2016). Although MCC can enable mobile devices to overcome resource shortages such as computing power, storage capacity, and energy, which remains some problems such as bandwidth constraints, unreliable links and latency when mobile devices access remote cloud services by using wireless signals or wireless networks. Therefore, MCC is not effective enough for delay-intensive applications such as high-quality video streaming, augmented reality (AR) and virtual reality (VR) (Tyng-Yeu & You-Jie, 2017).

DOI: 10.4018/IJITWE.297968 *Corresponding Author

In order to solve this problem, precursory researchers proposed the concept of mobile edge computing (MEC), which a key technology in the emerging fifth-generation network, which can host computing-intensive applications, and the network MEC is close to mobile users and provides context-aware services with the help of network information. MEC can support various applications that strictly require real-time response such as driverless vehicles, AR, VR, robotics, and immersive media by bringing cloudlets closer to mobile users.(Rahimi et al., 2020; Luo et al; 2019). Satyanarayanan et al., (2009) are the first to state that cloudlet is a new element to extend the cloud architecture of mobile devices and can access networks through high-speed wireless links such as Wi-Fi, and cloudlet is also called "data center in a box" and cloudlet technology is a supplement and extension to MCC. Ahuja & Rolli (2012) proposed that cloudlet, which is typically deployed at wireless access points (APs), has computing resources, reliable transmission and data processing ability, can process user task requests and reduce latency of user access to services. Therefore, compared with MCC, MEC is closer to mobile users than MCC, mobile devices can offload their computing tasks to the cloudlet or edge cloud by accessing the wireless network, which greatly reduces the access delay for mobile devices to access the cloud service and improve the task processing capability of the mobile device.

Most of the existing studies focus on user task scheduling on cloudlet (see, e.g., Mukherjee et al., 2019; Nayak et al., 2019; Zhang et al., 2018; Fei et al., 2018; Verbelen et al., 2014), cloudlet resource allocation(see, e.g., Chukhno et al., 2020; Wang et al., 2020; Josilo et al., 2020) and cloudlet task migration(see, e.g., Sun et al., 2019; Shen et al., 2019). However, little attention has been paid to cloudlet deployment in MEC. Due to the limited coverage of Wi-Fi, especially in highly computing environments with complex user distribution like wireless metropolitan area network (WMAN), it is possible to study how to use cloudlet to effectively handle the computationally intensive tasks offloaded by mobile devices, but it is also very essential to deploy cloudlet in such a complex environment. There are several cloudlet placement problems in networks that have been studied in recent years. The software-defined network (SDN) based Internet of Things(IoT) is applied to the problem of cloudlet placement, and the coexistence of APs in different Internet of Things is discussed (Zhao et al., 2018). The cloudlet placement problem that takes total energy consumption as the optimization goal while ensuring the quality of service(QoS) of users is proved to be an NP-hard problem and a decomposition algorithm based on the SDN framework is proposed to solve this problem(Yang et al., 2019). However, SDN networks are generally not suitable for networks operated by ISP and the access process of various IoT devices are highly simplified to treat resources as direct wireless connections, when it comes to the actual situation of multiple cloudlets, and this access process cannot be simplified to a direct wireless connection. Therefore, the problem of cloudlet placement in WMAN is considered (Zhang et al., 2019; Wei et al., 2020), in view of the relatively large scale of WMAN, the distribution of APs is considered and a normalized cut value is formulated to minimize the target WMAN segmentation model to minimize the average access delay from users to cloudlet (Liu, 2019). However, the cost of cloudlet deployment is not mentioned in the above literature. The budget of the cloudlet infrastructure service provider (ISP) is limited, from the perspective of the ISP, how to reduce the cost of cloudlet deployment is normally very important. Therefore, Mondal et al., (2019a) apply Karush-Kuhn-Tucker of Lagrangian function to optimize the deployment cost of cloudlet in fiber-wireless network. The simulated degradation algorithm is used to solve the problem of cost-aware cloudlet resource allocation (Raei et al., 2019). The long-term cost of cloudlet deployment and operation are considered (Mondal et al., 2019b). Fan et al., (2019) considered the cost of cloudlet deployment and the average end-to-end delay, and developed a Lagrangian heuristic algorithm to solve this problem. Wang et al., (2020) aimed to optimize the cost of cloudlet deployment and network delay, and proposed a fault-tolerant cloudlet deployment solution, and then a binary-based differential evolution cuckoo search algorithm was proposed to solve this problem.

However, these studies do not address the mobility of users and the heterogeneity of cloudlet in WMAN. Although there are few studies on minimizing the cost of heterogeneous cloudlet deployment in WMAN, it is very important and cannot be ignored. Because in a large-scale WMAN with a large

number of APs, which needs to deploy some APs for users offloading tasks, if a small number of cloudlets are deployed in dense areas, it will violate the QoS of users, otherwise, if a large number of cloudlets are deployed at sparsely populated areas, it will cause resource waste and increase the cost of cloudlet service providers. Therefore, Yao et al., (2017) first propose the minimum cost of heterogeneous cloudlet deployment while ensuring the QoS of users on MEC environment and formulate the problem as an integer linear programming (ILP). Because of the poor scalability of ILP, they propose a heuristic algorithm (HA), where each cloudlet with different capacities can be deployed all unoccupied APs, the HA select an optimal AP by Aps degree. However, the AP with a heavy workload may not be the closest to the user, which will increase user tolerable delay, meanwhile, it does not address the average latency of APs transmitting the user requests and the resource demands of the user task requests. Therefore, based on (Yao et al., 2017), this paper consider the mobility of users, the number of user task requests and the average delay of APs transmitting user task requests to improve and build a new heterogeneous cloudlet deployment cost model . The problem of minimizing the cost of heterogeneous cloudlet deployment can be divided into three sub-problems, including how many cloud servers are placed while ensuring users' QoS, choose which wireless APs are used for cloudlet deployment and how to place different capacity cloudlet servers according to different user densities in WMAN. The minimization the deployment cost of heterogeneous cloudlet servers in WMAN based on MEC environment is defined as an NP-hard problem, therefore, an improved heuristic algorithm (IHA) is proposed in this paper, which will combine the user request rate of each AP with the transmission delay between AP and cloudlet, calculate the average network latency to sort APs, and select an optimal location for cloudlet deployment.

Motivated by the above facts in this work and the contributions of this article can be summed up as follows:

- A new and more comprehensive cost-aware heterogeneous cloudlet deployment model is designed by introducing the number of user task requests and the average delay of APs transmitting user task requests. The heterogeneous cloudlet deployment model is designed to improve the QoS of end users and reduce the cost of cloudlet deployment.
- As against existing heuristic algorithm, the problem is formulated as an ILP. Accordingly, an IHA is developed, which combines the user request rate of each AP with the transmission delay between AP and cloudlet to select the optimal AP, so as to significantly reduce the delay and the cost of heterogeneous cloudlet deployment.
- Extensive experimentation and evaluation are conducted to verify the performance of the proposed algorithm, and the simulation results demonstrate that the IHA and designed model are more effective than HA.

The rest of this paper is organized as follows. Related work is introduced in Section 2. Section 3 presents model and problem formulation. Section 4 introduce the details of proposed algorithm. Section 5 displays experimental analysis. In the end, conclusion is shown in Section 6.

RELATED WORK

Most of the existing research focuses on cloudlet resource allocation, virtual machine migration and cloudlet deployment with time delay as the optimization goal based on MEC scenarios (Dolui et al., 2020). Mukherjee et al., (2019) studied how mobile users can select suitable cloudlet for task offloading in multiple cloudlet environments, the energy consumption and time delay are used as optimization goals, meanwhile, an optimal cloudlet selection strategy was proposed that can reduce power consumption and latency. Zhang et al., (2018) regarded the cloud task scheduling problem as a multiple direct acyclic graph scheduling problem, and the proposed scheduling strategy focused on the QoS of user resource demands and the cost of the cloudlet service providers. Fei et al., (2018)

proposed a multi-objective optimization model that considers security level, cloudlet access costs, and energy consumption. Sun et al., (2019) discussed the migration of cloudlet, of which objective is how to choose a suitable destination for cloudlet deployment. To select the optimal location of cloudlet placement, Shen et al., (2019) discussed the problem of minimizing the number of cloudlet deployments and proposed an energy-saving cloudlet migration method to effectively reduce the number of cloudlets. Yang et al., (2019) discussed the problem of cloudlet placement on the network and assign each requested task to cloudlets and public clouds to minimize the total energy consumption without violating the delay requirements of each task, a decomposition algorithm is proposed to solve the NP-hard problem. Verbelen et al., (2014) introduced a cloudlet architecture that is placed with wireless APs, and can also share resources between each other for cloudlet to offload. By adaptively configuring and outsourcing application components, a more fine-grained method is proposed to optimize the platform's applications based on mobile device functions and the available resources of cloudlet.

As an infrastructure, cloudlet can be deployed in different existing wireless network scenarios. Liu et al., (2019) aimed at the cloudlet placement model and optimization problem of WMAN, a cloudlet placement algorithm based on spectral clustering is designed. The algorithm takes into account the influence of factors such as the number of APs, the connection status between APs, the arrival rate of user requests of access points, and aims to optimize the access delay of mobile users offloading tasks to cloudlet, which has a good application prospect for the cloudlet of large-scale WMAN based on MEC. With the accelerating development of location-based services in mobile networks, Wei et al., (2020) proposed a service cache selection algorithm based on back-propagation neural network and users' mobility, and the proposed algorithm predicts the user's target location, the service request is thus forwarded to the appropriate target location through the service allocation algorithm to maximize the number of users of the local edge cloud service and reduce invalid service requests. Zhao et al., (2018) discussed cloudlet deployment for wireless optical networks and the Karush-Kuhn-Tucker is proposed to optimize this problem. However, it did not address the mobility of users and virtual machines. Zhang et al., (2019) studied the dynamic service placement of VR group games in a distributed MEC environment. In using the model predictive control framework to build online algorithms, and focused on designing approximate algorithms on each predictive window by solving a series of binary optimizations based on α-expanding through graphics-theoretical minimum shear to solve the problem and proved the performance guarantee of the boundary through this method. Mondal et al., (2019a) discussed the cloudlet deployment to support VR, and the service operation cost is used as optimal objective.

Although the issues regarding the location of the cloudlet deployment, user end-to-end delay, and user resource allocation have been well resolved, all the existing studies have adapted homogeneous cloudlet assumption and the cost of cloudlet deployment is ignored. The budget of infrastructure service provider is limited, it is a crucial issue for service providers to reduce the cost of cloudlet deployment. Accordingly, Raei et al., (2019) used simulated degradation algorithm to solve the problem of cost-aware cloudlet resource allocation and a mixed integer nonlinear programming model is proposed to optimize the cost of cloudlet deployment for static network planning. Fan et al., (2019) proposed cost-aware cloudlet placement strategy in the MEC, where cloudlet cost and average end-to-end latency are considered. A Lagrangian heuristic algorithm was developed to solve this problem. After placing the cloudlet on the network, a workload distribution scheme was designed by considering user mobility to minimize the E2E delay between the user and cloudlet. Wang et al., (2020) weighed the total network latency and the cost of cloudlet deployment in SDN-based IoT to minimize the total cost of the cloudlet network, and a fault-tolerant cloudlet deployment scheme is proposed, and then, a binary-based differential evolution cuckoo search algorithm is developed to optimize the cost of cloudlet deployment and network delay. Mondal et al., (2019b) focused on the static cloudlet network planning problem, and proposed a hybrid cost optimization framework for the optimal placement for the existing passive optical access network, and develop a mixed integer

nonlinear procedure to determine the cloudlet placement location. However, these studies do not address the mobility of users and the heterogeneity of cloudlets in WMAN.

Considering the heterogeneity of cloudlets in the IoT environment, Yao et al., (2017) used a low-complexity heuristic algorithm to study how to deploy servers in a cost-effective way without violating the predetermined quality of service, on the one hand, it do not address the average delay of APs transmitting user requests and the resource requirements of user task requests, on the other hand, the method of selecting an optimal AP for cloudlet deployment is sorted according to the degree of wireless APs, however, APs with heavier workloads are not necessarily the closest to the users they serve, which will result in higher user tolerance latency. Therefore, in this paper, the contact probability of users with the wireless AP and the transmission delay between the user offloading task request to the cloudlet are considered to calculate the average network delay of the APs, and sort the APs by the average network delay. From all the above literature, the authors notice that most of existing studies are with homogeneous cloudlet assumption in WMAN, and the existing cost-aware heterogeneous cloudlet deployment models need to be improved and suitable for small network areas. Therefore, this paper studies how to provide heterogeneous cloudlet placement strategies with different service levels according to the different resource demands of users in a WMAN scenario.

MODEL AND PROBLEM FORMULATION

System Model

As shown in Figure 1, the entire system consists of four roles, including APs, cloudlets, users, and the set of links. A WMAN is thus defined as a connected and undirected graph $G = \left\{V \cup S \cup U,\ E\right\}$, where $V = \left\{v_1,\ v_2,\ v_3,\ \ldots\ ,v_m\right\}$ represents *m* APs in WMAN, each AP covers an area with other APs and can communicate with other APs directly or through multi-hop in (Liu, 2019; Dolui et al., 2020). For users, $U = \left\{u_1,\ u_2,\ u_3,\ \ldots\ ,u_n\right\}$ denotes the set of *n* mobile users which randomly roam in the area covered by APs, and mobile users have time-varying locations and resource demands towards APs located at cloudlets, which lead to different request rates for APs connected to the user. Accordingly, the user task request of each AP may be unpredictable, especially when the user moves within a period of time. Therefore, it is assumed that each AP point has a task flow that can be offloaded and arrives at the system randomly and obeys Poisson distribution in (Liu; 2019), meanwhile, it is assumed that the user request rate of AP v_i is ρ_i, which can be can be accurately estimated by fitting method in (Luo et al; 2019).

Therefore, it is assumed that R_k at each AP v_k represents the number of user requests. Let the number of tasks that AP v_k has received from the user u_i be N_{ik}, as shown in Figure 2, $N_{11} = R_1 * p_{11}$ denotes the number of tasks that AP v_1 has received from the user u_1, accordingly, $N_{1m} = R_1 * p_{m1}$ denotes the number of tasks that AP has received from u_m. The number of task requests of all users that associate with AP v_j received by v_j can be captured as Equation (1):

$$N(v_j) = \sum_{i \in U} R_i * p_{ij}\ ,\ \forall j \in V \tag{1}$$

Figure 1. Cloudlet deployment for MEC

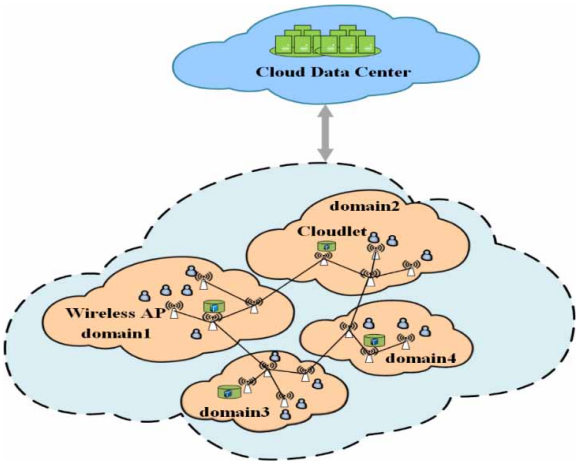
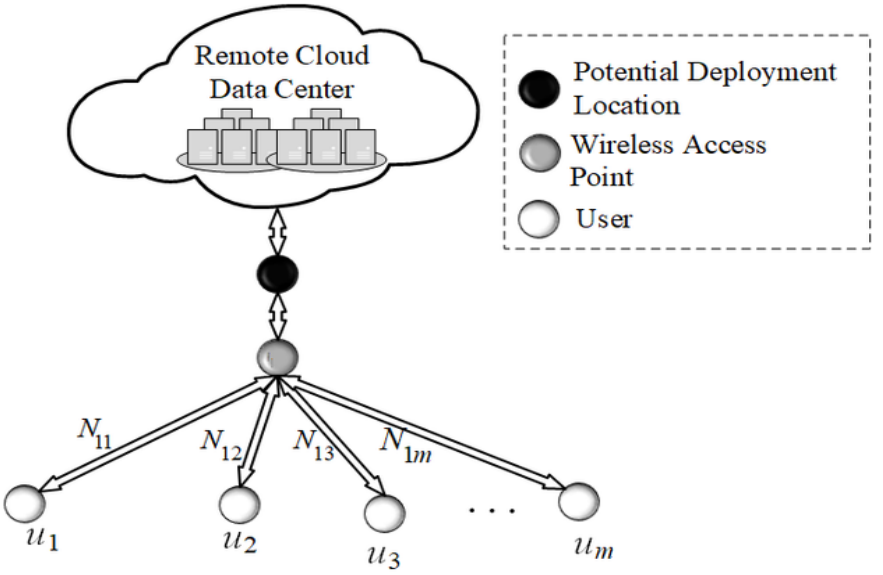


Figure 2. An example of the number of user tasks received by AP



S represents a group of potential locations of cloudlets and *E* denotes each link between two APs in *V* or between an AP and a potential location in *S*. $F = \{f_1, f_2, f_3, \ldots, f_k\}$, $1 \le k \le |S|$ represents the set of cloudlet servers. In order to reduce the transmission latency between mobile devices and the remote cloud, the ideal location for cloudlet shall be a network location that is one hop away from the mobile device such as cellular base station or Wi-Fi AP. Accordingly, It is assumed that the deployment location of the cloudlet is the same as APs, and *k* cloudlets need to be deployed to *k* different potential locations in the set *S*. Different users have different resource demands, dm_i refers to user resource demands for user $u_i \in U$, the user's total resource demand shall not exceed the resource capacity provided by the server f_i. The deployment cost and resource capacity of the server

are denoted by W_k and r_k respectively. It is assumed that the servers are heterogeneous, $f_i \neq f_j, W_i \neq W_j$, and different cloudlet servers have different costs and resource capacities in (Yao et al; 2017).

For each link $\left(v_i,\ v_j\right)$ in *E*, define the latency of transmitting a user request between two endpoints (APs) v_i and v_j as the shortest path value between the two points, d_{ij} denotes the latency of transmitting user requests between v_i and a cloudlet located at AP v_j. When the user u_i request is transmitted to the nearest AP v_i through the wireless network, the request delay can be considered as 0, otherwise, the user u_i request is transmitted to a cloudlet AP f_j deployed at v_j in a multi-hop manner, the transmission latency cannot be ignored. The definitions of the main symbols used in this paper are shown in Table 1.

Table 1. Symbol Definition

Symbols	Definition
$G = \left\{V \cup S \cup U,\ E\right\}$	APs set, potential cloudlet locations set and the mobile users set.
$m = \left\|V\right\|,\ \in = \left\|E\right\|,\ n = \left\|U\right\|$	The number of APs in *V*, the number of links in *E*, and the number of users in *U*.
R_j	User request collection of AP v_j.
ρ_i	User request rate in AP v_i.
p_{ik}	Contact probability between user u_i and AP v_k.
N_{ik}	The number of tasks that AP v_k receives from user u_i.
dm_i	The resource demand of user u_i.
$d(e)$	Link delay between APs.
T_r	Delay tolerance of user u_r.
d_{ij}	Transmission delay between AP v_j and v_i.
D_{ij}	The latency of user u_i offloading tasks to a cloudlet located at v_j.
D_j	The average delay of AP v_j transmitting user requests.
W_k	The deployment cost of cloudlet server f_k.
r_k	Resource capacity of cloudlet server f_k.
P_{tol}	The total cost of cloudlet servers.

Problem Statement

The key to the problem is how to place the cloudlet server to minimize the deployment cost of *k* heterogeneous cloudlet servers. Meanwhile, the average delay for each AP to transmit user requests shall not exceed the tolerable delay for users. The total resource demands for users' request transmitted by each AP does not exceed the provisioned resource capacity. The cost-aware heterogeneous cloudlet deployment problem can be mathematically described as follows.

The cost-aware heterogeneous cloudlet deployment problem can be formulated as an ILP. For $v_j \in [1,\ k]$ and $v_i \in [1,\ |S|]$, where $\beta_{ij} = 1$ if cloudlet f_j is deployed at AP v_i, $\beta_{ij} = 0$ otherwise. $\varphi_{ij} = 1$, if the task request of user u_i is offloaded to a cloudlet located at v_j and $\varphi_{ij} = 0$, otherwise. The number of tasks of user u_i received by AP v_k is closely related to the contact probability q_{ik} between the user u_i and AP v_k, N_{ik} is calculated as Equation (2):

$$N_{ik} = R_k * p_{ik},\ \forall i \in U,\ \forall k \in V \tag{2}$$

Where

$$\sum_{k \in V} p_{ik} = 1,\ \forall i \in U$$

Therefore, the number of task requests of all users that associate with AP v_j received by v_j $\sum_{i \in U} N_{ij}$. D_{ij} represents the delay for user u_i offloading to the cloudlet located at v_i, which is expressed as Equation (3):

$$D_{ij} = \sum_{k \in V} N_{ik} * d_{kj} * \varphi_{ij},\ \forall j \in S,\ \ \forall i \in U \tag{3}$$

The average delay of AP v_j transmitting user requests be expressed as Equation (4):

$$D_j = \frac{\sum_{i \in U} \varphi_{ij} * D_{ij}}{\sum_{i \in U} N_{ij}},\ \forall j \in S \tag{4}$$

The objective of cost-aware heterogeneous cloudlet deployment problem is to minimize the cost of cloudlet deployment, which is described as Equation (5):

$$\text{Minimize: } P_{tol} = \sum_{j \in V} \sum_{k \in F} W_k * \beta_{jk} \tag{5}$$

subject to the following constraints:

$$\varphi_{ij} = \begin{cases} 1, \textit{ if task request of user } u_i \textit{ is offloaded to a cloudlet located at AP } v_j \\ 0, \textit{ otherwise} \end{cases} \tag{6}$$

$$\beta_{ij} = \begin{cases} 1, \text{ if cloudlet server } f_k \text{ is deployed to AP } v_i \\ 0, \text{ otherwise} \end{cases} \tag{7}$$

$$\sum_{j=1}^{|S|} \beta_{ij} = 1, \quad \forall 1 \le \mathrm{i} \le K \tag{8}$$

$$\sum_{i=1}^{|F|} \beta_{ij} = 1, \quad \forall j \in S \tag{9}$$

$$\sum_{i \in U} \varphi_{ij} * p_{ik} * dm_i \le \sum_{k \in F} r_k * \beta_{jk}, \quad \forall j \in S \tag{10}$$

$$\frac{\sum_{i \in U} \varphi_{ij} * D_{ij}}{\sum_{i \in U} N_{ij}} \le T_i \ , \quad \forall j \in S \tag{11}$$

Where constraint(8) ensures that each of the *k* cloudlet servers can only be deployed to one potential location from the set *S,* and constraint(9) ensures that each potential access point in set *V* should deploy one cloudlet server selected from set *F*. In order to avoid resource overload of cloudlet servers, constraint (10) ensures that the total resource demand from related users cannot exceed the provisioned resource capacity. To ensure users' QoS, depending on the cloudlet relationship, constraint (11) ensures that the average delay for each AP transmit user requests to cloudlet does not exceed the given user delay tolerance.

ALGORITHM DESIGN

Based on (Yao et al; 2017), the proposed algorithm combines the user request rate of each AP with the transmission delay between AP and cloudlet, sorts APs by calculating the average network latency, and selects an optimal location for cloudlet deployment. In this paper, the problem of minimizing the cost of heterogeneous cloudlet deployment in WMAN is divided into three sub-questions, including cloudlet server selection (lines 1-4 of Algorithm 1), cloudlet server deployment(lines 5-11 of Algorithm 1) and the QoS of users(lines12-27 of Algorithm 1).

For the server selection problem, the greedy strategy with the smallest unit resource cost is adopted, regarding the question of how many servers to select, user mobility and contact probability p_{ik} can be taken into to select resource capacity of cloudlet servers that needs to meet the total resource demand generated by related users contacting the AP $\sum_{i \in U} \varphi_{ij} * p_{ik} * dm_i$. Therefore, first select a resource capacity of a cloudlet server is greater than the total resource demand generated by users contacted by AP (line 2of Algorithm 1).

For the cloudlet server deployment, aiming at the problem of selecting an optimal AP, the mobility of mobile users is taken into account in this paper. Because of different user request arrival rate of APs and the shortest data transmission delay between APs, it is necessary to combine the user request arrival rate of APs ρ_i and the shortest data transmission delay between APs d_{ij} to calculate the average access delay of each AP in (Liu, 2019), which can balance the workload of APs, user density and

transmission cost between APs. Then sort the APs according to their average access delay to determine an optimal AP for cloudlet deployment (line 6 of Algorithm 1), *m* is the number of APs connected to AP v_j. The method of selecting an optimal AP is formulated as Equation (12):

$$A_j = \frac{\sum_{i=1}^{m} \rho_i d_{ij}}{m} \quad \forall j \in V \tag{12}$$

To ensure QoS of users, the average delay for each AP to transmit user requests to cloudlet does not exceed the given user delay tolerance, $D_j \leq ar_j$ (line 15 of Algorithm 1).The total resource demand from related users cannot exceed the resource capacity provided by the server $\sum_{i \in U} \varphi_{ij} * p_{ik} * dm_i \leq capacity_i$ (line 16 of Algorithm 1), a two-layer loop is adopted to select the cloudlet server with the lowest deployment cost from the candidate subset (lines 8-23 of Algorithm 1). If both $D_j \leq ar_j$ and $\sum_{i \in U} \varphi_{ij} * p_{ik} * dm_i \leq capacity_i$ are satisfied, the subset is marked as the final choice. After exiting the loop, the number of servers with the lowest cost is obtained.

Algorithm 1

Computational Complexity Analysis

It can be seen from Algorithm 1 that the maximum number of iterations of IHA is similar to that in (Yao et al., 2017), of which number of iterations is O ($|subset| * |s_i| * N$), subset represents the candidate subset of all servers. A set of candidate subsets of cloudlet servers can be defined as s_i. $|s_i|$ cannot exceed m APs, and a subset of candidate servers can be obtained in advance. Therefore, Algorithm 1 has polynomial time complexity.

Table 2. *Improved heuristic algorithm (IHA)*

Improved Heuristic Algorithm(IHA)
Input: $G=\{V \cup S \cup U,\ E\}$, r_k, R_k, d_{ij}, dm_i, T_r, W_k, ρ_i
Output: P_{tol}
1: /Cloudlet Server Selection/
2: subset Find all server subsets that meets the user's demands
3: Sort subset in increasing order of cost
4: Sort subset in decreasing order of resource capacity
5: /Cloudlet Server Deployment/
6: L_{AP} Sort AP in descending order by A_j
7: P_{tol} =max

Table 2 continued on next page

Table 2 continued

Improved Heuristic Algorithm(IHA)
8: for all $s_i \in subset$ do
9: $\varphi'_{ij} = \{0\}; \beta'_{ij} = \{0\}; P'_{tol} = 0;$
10: for all $k \in S_i, j \in L_{AP}$ do
11: $\beta'_{jk} = 1 \; ; \; P_{tol} += W_k$
12: /The QoS of Users/
13: for all $i \in U$ do
14: $ar_i = T_r; \; flag_i = 0; D_j = \frac{\sum_{i \in U} \varphi_{ij} * D_{ij}}{\sum_{i \in U} N_{ij}}$
15: if $D_j \leq ar_j$ then
16: if $\sum_{i \in U} \varphi_{ij} * p_{ik} * dm_i \leq capacity_i$ then
17: end if
18: end if
19: if $flag_i == 1$ then
20: Update the resource capacity of v_j
21: end if
22: end for
23: end for
24: if $P_{tol} \leq P'_{tol}$ then
25: $P_{tol} = P'_{tol}; \; \varphi_{ij} = \varphi'_{ij};$
26: end if
27: end for

EXPERIMENT AND ANALYSIS

In the existing research, there is no algorithm to directly solve this problem, the authors compare the performance between the IHA and HA from four aspects under different network scales, including the number of users, the number of servers, the maximum resource demand of users, the maximum resource capacity of cloudlet servers. The authors are the first to design a complete cost-aware heterogeneous cloudlet server model, which comprehensively consist of user mobility, cloudlet heterogeneity, number of user requests, and the average delay for AP to transmit user requests is calculated to ensure user QoS.

Experimental Settings

All the experimental settings are the same as those in (Liu, 2019; Yao et al., 2017). Barabasi-Albert model in the Networkx package in Python3.7 is used to generate the random network $G=(V \cup S \cup U, E)$, each direct link d(e) is generated in [5ms, 50ms]. To construct a network transmission delay matrix between APs d_{ij}, the Floyd algorithm is used to calculate the shortest delay between each pair of wireless APs. The parameter values in the simulations are set as Table 3. It can be seen that these parameters are adjustable and scalable. The number of cloudlet servers is half of the number of wireless APs. We first studied the performance and scalability of the improved algorithm on solving the new model by transforming the number of cloudlet servers from 5 to 50 and the number of users from 5 to 50. Then, we evaluate the performance of improved algorithm by varying cloudlet capacities from 10 to 80 and tolerable service access delay. HA-MUAD is used to represent the minimum user access delay by HA, IHA-AMURD is used to represent the average minimum user request delay solved by IHA, P(AMURD /MUAD) represents the percentage of the average minimum user request delay by IHA is less than the minimum user access delay by HA, and P(HA- P_{tol} / IHA- P_{tol}) represents the percentage of the deployment cost P_{tol} by HA that is less than the deployment cost P_{tol} by IHA.

Table 3. The experimental parameter settings

Notations	Parameter Settings	Values
Resource capacity of a server f_k.	r_k	[10, 500]
Resource demand of a user u_i.	dm_i	[1, 50]
Delay tolerance of user u_r (ms).	T_r	[10, 500]
User request collection of AP v_k.	R_k	[50, 500]
Link delay between APs.	$d(e)$	[5, 50]

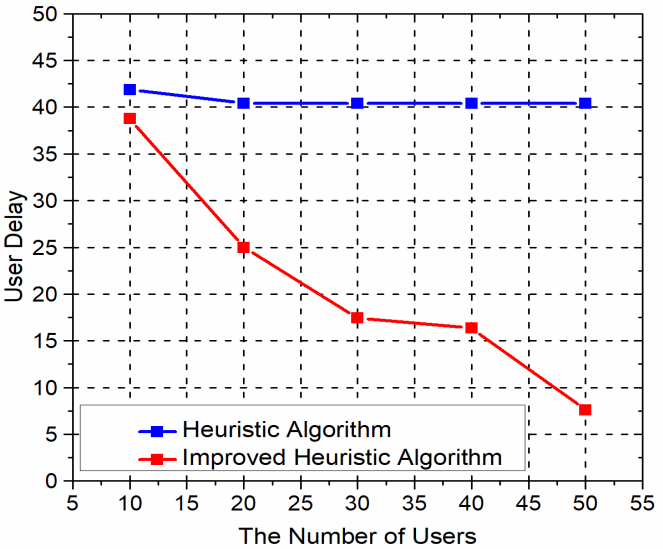
Effect of Number of Users On Delay and Cost

In this section, the number of cloudlet servers $server(k)$ is set to 10 and the number of APs is set to 20. The authors analyze the effect of the number of users on user delay by setting the number of users $user(n) \in [10,\ 50]$. As shown in Table 4 and Figure 3, when the value of $user(n)$ is 10, 20,30,40 and50 respectively, the IHA-AMURD is 7.47%, 38.21%, 56.86%, 59.56% and 81.20% less than HA-MUAD respectively. The lower limit of T_i of IHA is thus less than the lower limit of T_i of HA. Because user mobility, cloudlet heterogeneity, number of user requests, and the average delay for AP to transmit user requests are comprehensively considered to ensure users' QoS. Consequently, when the number of users increases from 10 to 50, the minimum average delay for each AP transmitting user requests gradually decreases. Therefore, as against HA, the new model and the IHA are close to the optimal solution for different numbers of users, of which the minimum average delay of AP transmitting user requests is relatively low.

Table 4. The relationship between the number of users and delay

Variables	HA		IHA		Percentage
$user(n)$	MUAD (ms)	T_i (ms)	AMURD (ms)	T_i (ms)	P(AMURD /MUAD)
10	41.89	[60, 100]	38.76	[50, 100]	7.47%
20	40.43	[60, 100]	24.98	[40, 100]	38.21%
30	40.43	[60, 100]	17.44	[30, 100]	56.86%
40	40.43	[60, 100]	16.35	[30, 100]	59.56%
50	40.43	[60, 100]	7.60	[30, 100]	81.20%

Figure 3. Delay on different number of users

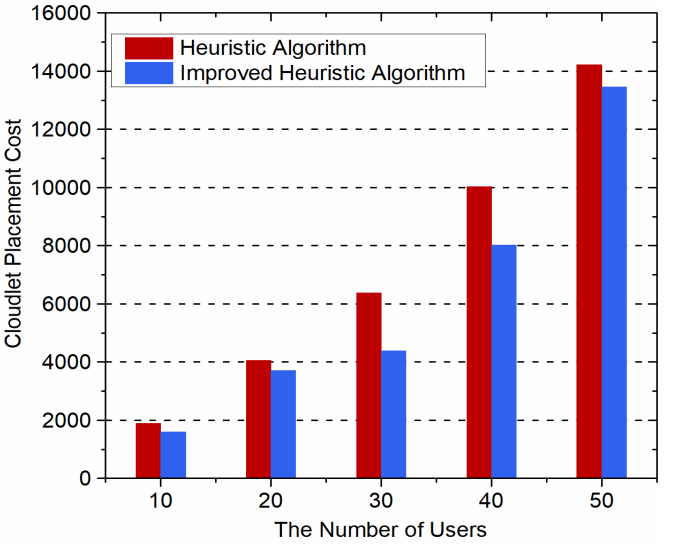


It can be seen from Table 5 and Figure 4 that the number of users $user(n)$ increased from 10 to 50, and the cost by IHA was lower than the cost by HA, indicating that the better performance of IHA and the new model. when the value of $user(n)$ is 10, 20,30,40 and50 respectively, the deployment cost by IHA is 12.40%, 8.25%, 31.32%, 19.95% and 5.30% less than the cost by HA respectively. Meanwhile, compared with HA, IHA only needs less resource capacity r_k to meet the user's task requirements. Therefore, the lower and upper limit of r_k of IHA is less than the lower limit of r_k of HA. Although when $user(n)$ =50, $r_k \in [300, 400]$ in IHA, since the designed model considers the contact probability between users and wireless APs, the deployment cost by IHA was lower than the cost by HA. The above experimental results show that IHA and designed model is more effective than HA.

Table 5. The relationship between the number of users and cost

Variables	HA		IHA		Percentage
$user(n)$	r_k	P_{tol}	r_k	P_{tol}	P(HA- P_{tol} / IHA- P_{tol})
10	[50, 150]	1814	[30, 130]	1589	12.40%
20	[100, 200]	4049	[50, 150]	3715	8.25%
30	[150, 250]	6385	[100, 200]	4385	31.32%
40	[200, 300]	10027	[170, 270]	8027	19.95%
50	[250, 350]	14215	[300, 400]	13462	5.30%

Figure 4. Deployment cost on different number of users



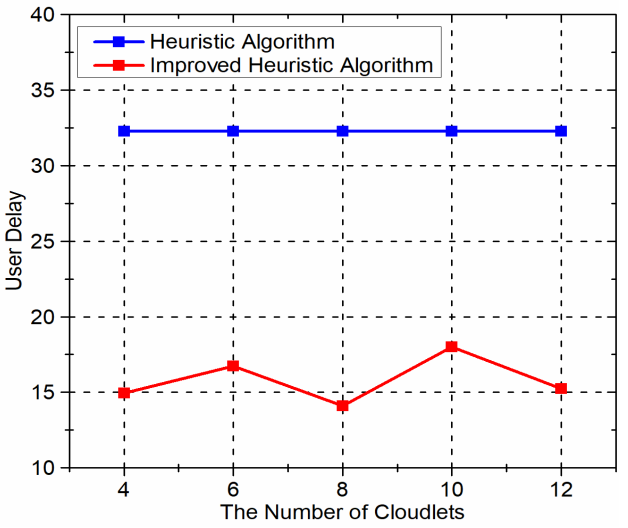
Effect of Number of cloudlets On Delay and Cost

In this section, the $user(n)$ and APs are set to 20 and 15 respectively. As shown in Table 6, adjust $server(k)$ from 4 to 12. In a group of experiments, the cloudlet servers cannot meet the user's resource demands when the number of cloudlet servers is less than 4, and no feasible solution can be found. Therefore, the number of cloudlet servers is set to be greater than or equal to 4. Similarly, it can be seen from Figure 5 that the number of cloudlet servers increases from 4 to 12, the IHA-AMURD gradually decreases. Because as the the number of cloudlets increases, the cloudlet server is already sufficient and stable to meet users' resource demands, and there is no need to further include more candidate servers.

Table 6. The relationship between the number of cloudlets and delay

Variables	HA		IHA		Percentage
$server(k)$	MUAD (ms)	T_i (ms)	AMURD (ms)	T_i (ms)	P(AMURD /MUAD)
4	32.28	[50, 100]	14.96	[30, 100]	53.66%
6	32.28	[50, 100]	16.73	[30, 100]	48.17%
8	32.28	[50, 100]	14.12	[30, 100]	56.26%
10	32.28	[50, 100]	17.99	[30, 100]	44.27%
12	32.28	[50, 100]	15.24	[30, 100]	52.79%

Figure 5. Delay on different number of cloudlets



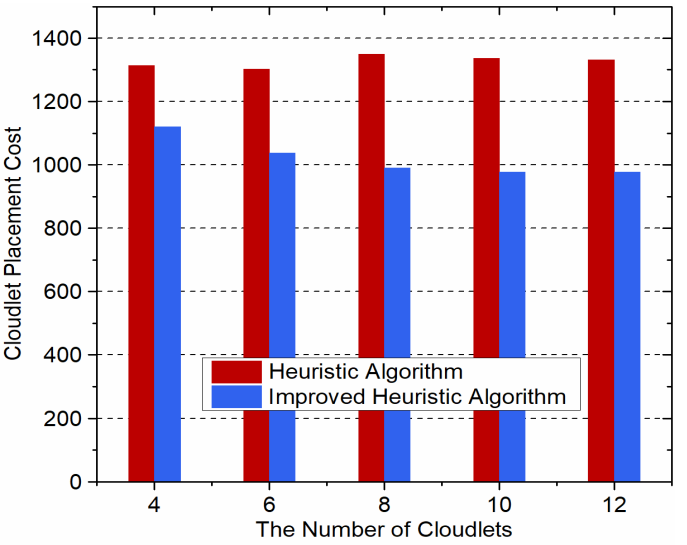
It can be seen from Table 7 and Figure 6 that the number of cloudlet servers $server(k)$ has increased from 4 to 12. The lower and upper limit of r_k of IHA is greater than the lower and upper limit of T_i of HA, however, when the value of $server(k)$ is 4, 6, 8, 10 and12 respectively, the cost

by IHA is 14.70%, 20.41%, 26.59%, 26.80% and 26.52% less than the cost by HA respectively. The resource capacity of each cloudlet server is proportional to the cost of the cloudlet server, meanwhile, when the number of cloudlet servers increases from 4 to 12, the deployment cost by IHA gradually decreases. Therefore, it can be proved that the higher performance of new model and the improved IHA, which can optimize the resource capacity of the cloudlet server while reducing the total cost of the cloudlet server.

Table 7. The relationship between the number of cloudlets and cost

Variables	HA		IHA		Percentage
$server(k)$	r_k	P_{tol}	r_k	P_{tol}	P(HA- P_{tol} / IHA- P_{tol})
4	[100,200]	1313	[150,250]	1120	14.70%
6	[100,200]	1303	[150,250]	1037	20.41%
8	[100,200]	1350	[150,250]	991	26.59%
10	[100,200]	1336	[150,250]	978	26.80%
12	[100,200]	1331	[150,250]	978	26.52%

Figure 6. Deployment cost on different number of cloudlets



Effect of the Maximum Resource Capacity of Cloudlets on Delay and Cost

The number of users, APs and cloudlet servers are set to 50, 20, and 10 respectively. As shown in Table 8 and Figure 7, with larger resource capacity, less cloudlet servers should be deployed to satisfy users' task requirements. when the value of r_k is 200, 250,300,350 and 400 respectively, the cost by IHA is 76.30%, 67.25%, 78.60%, 72.08% and 76.30% less than the cost by HA respectively. Meanwhile,

the upper and lower limit of $T_i \in [20,50]$ of IHA is lower than the upper and lower limit of $T_i \in [50,\ 500]$ of HA. Therefore, the new model and the improved algorithm can be effectively applied to reduce the user tolerance delay while ensuring users' QoS.

Table 8. The relationship between the maximum resource capacity of cloudlets and delay

Variables	HA		IHA		Percentage
r_k	MUAD (ms)	T_i (ms)	AMURD (ms)	T_i (ms)	P(AMURD/MUAD)
200	40.43	[50, 500]	9.58	[20, 50]	76.30%
250	40.43	[50, 500]	13.24	[20, 50]	67.25%
300	40.43	[50, 500]	8.65	[20, 50]	78.60%
350	40.43	[50, 500]	11.29	[20, 50]	72.08%
400	40.43	[50, 500]	9.58	[20, 50]	76.30%
450	40.43	[50, 500]	8.66	[20, 50]	78.58%
500	40.43	[50, 500]	8.66	[20, 50]	78.58%

As shown in Table 9 and Figure 8, for HA, when $user(n)$ is 50 and r_k is 200, While ensuring the user QoS under the same capacity of the cloudlet server, the cloudlet servers handling more user resource demands can reduce cost and task waiting latency in a certain, therefore, the deployment cost shows as a decreasing function. For HA, the feasible and optimal solution can be found when the user resource demand $dm_i \in [1,10]$. However, for IHA, $dm_i \in [1,15]$, the feasible and optimal solutions can be found. Similarly, the upper limit of dm_i of IHA is greater than that of HA when r_k increases from 200 to 500. Nevertheless, the high efficiency of the proposed Algorithm by it outperforms the HA in solving user delay and deployment cost.

Figure 7. Delay on different maximum resource capacity of cloudlets

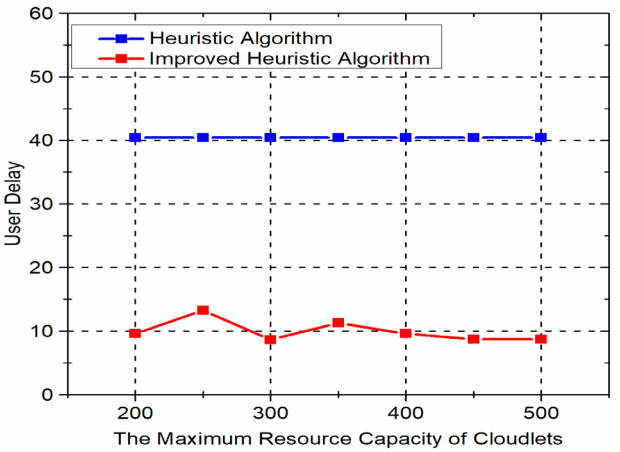
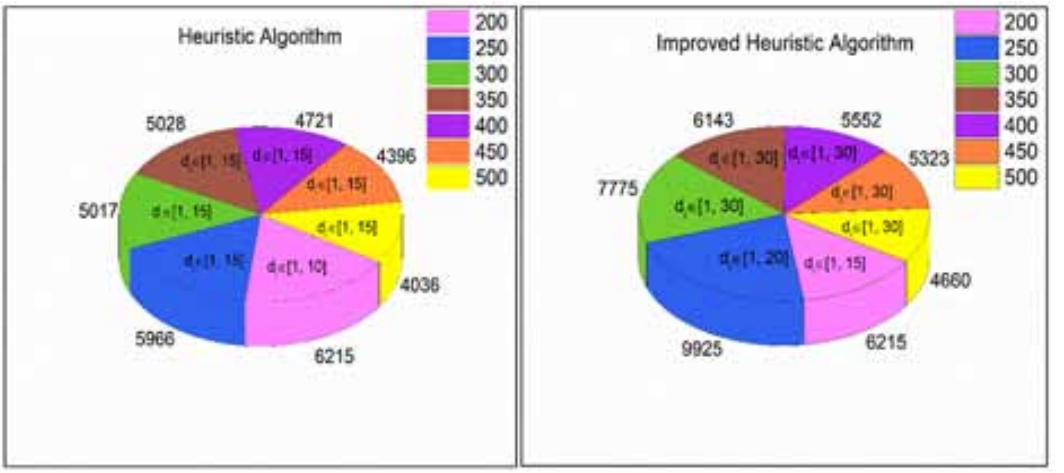


Table 9. The relationship between the maximum resource capacity of cloudlets and cost

Variables	HA		IHA	
r_k	dm_i	P_{tol}	dm_i	P_{tol}
200	[1, 10]	6215	[1, 15]	6215
250	[1, 15]	5966	[1, 20]	9925
300	[1, 15]	5017	[1, 30]	7775
350	[1, 15]	5028	[1, 30]	6143
400	[1, 15]	4721	[1, 30]	5552
450	[1, 15]	4396	[1, 30]	5323
500	[1, 15]	4036	[1, 30]	4660

Figure 8. Deployment cost on different maximum resource capacity of cloudlets



Effect of the Maximum Resource Demand of Users on Cost and Delay

As shown in Table 10 and Figure 9, this chapter evaluates the impact of the maximum resource demands of users on cost and delay. The fixed minimum user resource demand dm_i is 10, and the number of $user(n)$, APs and $server(k)$ is 50, 20 and 10 respectively. When the maximum resource demand is increased from 10 to 80, the IHA-AMURD gradually decreases, which is lower than HA-MUAD. Consequently, the experimental results show that IHA is more effective than HA.

As shown in Table 11 and Figure 10, when the resource capacity of users is 20 and 40 respectively, there is no difference between the deployment cost in HA and IHA respectively. Therefore, H_0 hypothesis is used to analyze the difference between the deployment cost of cloudlet servers of HA and IHA, and then, one-way Analysis of Variance (ANOVA) is applied to analyze the efficiency of the IHA algorithm by calculating the difference between the cost by IHA and the cost by HA. In

one-way ANOVA, the significance level α is set to 0.1, after calculating, P-value is 0.993019. Mathematically, H0 hypothesis $\mu_1 = \mu_2$, the alternative hypothesis H1: $\mu_1 \neq \mu_2$. Since p-value: 0.993019 > α: 0.1, H_0 is accepted, which demonstrates the difference between the deployment cost of HA and IHA is not big enough to be statistically significant. However, the deployment cost of cloudlet subject to the resource demand of users and resource capacity of cloudlet servers. The value range of the resource capacity of the IHA server is not only smaller than the value range of the HA resource capacity, but also can meet user requests with a lower deployment cost. Overall, IHA and the improved model have higher performance than those of HA.

Table 10. The relationship between the maximum user resource demand and delay

Variables	HA		IHA		Percentage
dm_i	**MUAD (ms)**	T_i (ms)	AMURD **(ms)**	T_i (ms)	P(AMURD / MUAD)
10	40.43	[50, 500]	10.87	[20, 50]	73.11%
20	40.43	[50, 500]	10.32	[20, 50]	74.47%
30	40.43	[50, 500]	11.35	[20, 50]	71.93%
40	40.43	[50, 500]	10.32	[20, 50]	74.47%
50	40.43	[50, 500]	7.60	[20, 50]	81.20%
60	40.43	[50, 500]	11.35	[20, 50]	71.93%
70	40.43	[50, 500]	10.87	[20, 50]	73.11%
80	40.43	[50, 500]	13.24	[20, 50]	67.25%

Figure 9. Delay on different maximum resource demand of users

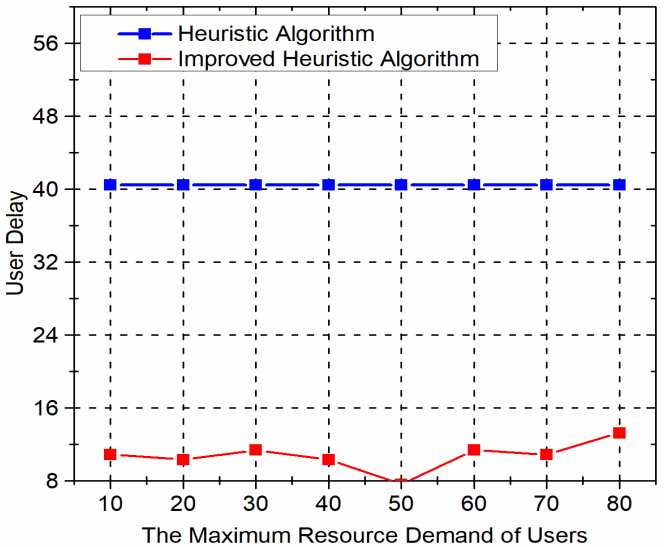
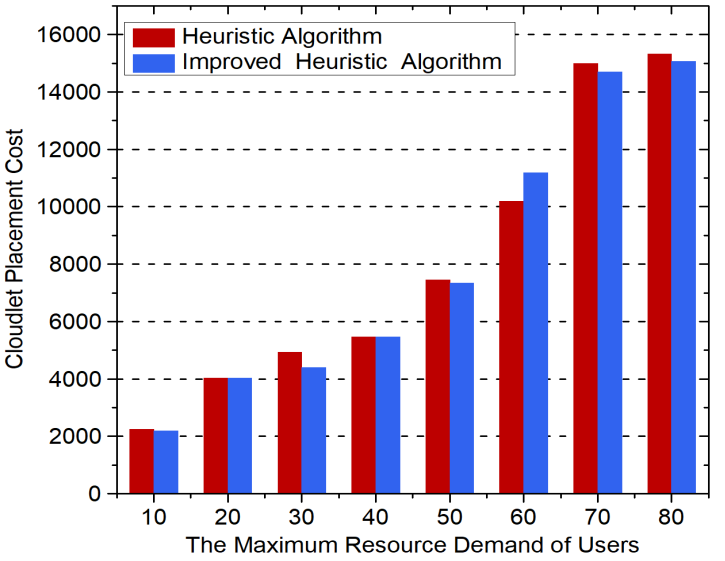


Table 11. The relationship between the maximum resource demand of users and cost

Variables	HA		IHA	
dm_i	r_k	P_{tol}	r_k	P_{tol}
10	[10, 200]	2244	[10, 200]	2201
20	[10, 250]	4038	[10, 230]	4038
30	[10, 300]	4931	[10, 250]	4402
40	[50, 500]	5464	[50, 500]	5464
50	[50, 500]	7454	[50, 500]	7735
60	[50, 500]	10184	[50, 500]	11192
70	[450, 500]	14982	[50, 500]	14691
80	[450, 500]	15316	[50, 500]	15067

Figure 10. Deployment cost on different maximum resource demand of users



CONCLUSION

In this paper, the authors design a new and more comprehensive cost-aware heterogeneous cloudlet deployment model by introducing the number of user task requests and the average delay of APs transmitting user task requests, which is designed to improve the QoS of end users and reduce the cost of cloudlet deployment. Meanwhile, the authors develop the IHA with the method of selecting an optimal AP for cloudlet deployment and ensuring the QoS of users, the latency and the cost of heterogeneous cloudlet deployment are significantly reduced. The experimental results verify the

high efficiency and high performance of the model designed and the improved heuristic algorithm. In the future, we will optimize the deployment cost and network delay of cloudlet in WMAN, and use the number of servers deployed by wireless AP nodes and user resource capacity as constraints to optimize the heterogeneous cloudlet deployment model.

FUNDING INFORMATION

The publisher has waived the Open Access Processing fee for this article.

ACKNOWLEDGMENT

This research was supported by the National Natural Science Foundation of China [grant number 61802085], the Guangxi Natural Science Foundation [grant number 2020GXNSFAA159038], the Foundation of Guilin University of Technology [grant number GUTQDJJ2002018], and the Guangxi Universities key Laboratory Director Fund of Embedded Technology and Intelligent Information Processing [grant number 2020-1-7].

REFERENCES

Ahuja, S., & Rolli, A. (2012). Exploring the convergence of mobile computing with cloud computing. *Network and Communication Technologies*, *1*(1), 1–97. doi:10.5539/nct.v1n1p97

Chukhno, O., Chukhno, N., Araniti, G., Campolo, C., Iera, A., & Molinaro, A. (2020). Optimal placement of social digital twins in edge IoT networks. *Sensors (Basel)*, *20*(3), 6181–6199. doi:10.3390/s20216181 PMID:33143038

Dolui, K., & Datta, S. K. (2017). Comparison of edge computing implementations: Fog computing, cloudlet and mobile edge computing. In 2017 Global Internet of Things Summit (GIoTS) (pp. 1-6). IEEE. doi:10.1109/GIOTS.2017.8016213

Fan, Q., & Ansari, N. (2019). On cost aware cloudlet placement for mobile edge computing. IEEE/CAA. *Journal of Automatica Sinica*, *6*(4), 926–937. doi:10.1109/JAS.2019.1911564

Fei, H., Doo-Soon, P., Jungho, K., & Geyong, M. (2018). 2L-mc: A two-layer multi-community-cloud/cloudlet social collaborative paradigm for mobile edge computing. *IEEE Internet of Things Journal*, *6*(3), 4764–4773. doi:10.1109/JIOT.2018.2867351

Gai, K. K., Qiu, M. K., Zhao, H., Tao, L. X., & Zong, Z. L. (2016). Dynamic energy-aware cloudlet-based mobile cloud computing model for green computing. *Journal of Network and Computer Applications*, *59*, 46–54. doi:10.1016/j.jnca.2015.05.016

Josilo, S. (2020). *Task Placement and Resource Allocation in Edge Computing*. Retrieved from https://www.diva-portal.org

Liu, Y. P. (2019). *Research of cloudlet placement strategy based on spectral clustering in Mobile Edge Computing* (Unpublished master's thesis). Southwest University, Chongqing, China.

Luo, Y., & Qiu, S. (2019). *Optimal resource reservation scheme for maximizing profit of service providers in edge computing federation. In 2019 IEEE Intl Conf on Dependable, Autonomic and Secure Computing, Intl Conf on Pervasive Intelligence and Computing, Intl Conf on Cloud and Big Data Computing, Intl Conf on Cyber Science and Technology Congress (DASC/PiCom/CBDCom/CyberSciTech)*. IEEE. doi:10.1109/DASC/PiCom/CBDCom/CyberSciTech.2019.00159

Mondal, S., Das, G., & Wong, E. (2019a). Efficient cost-optimization frameworks for hybrid cloudlet placement over fiber-wireless networks. *Journal of Optical Communications and Networking*, *11*(8), 437–451. doi:10.1364/JOCN.11.000437

Mondal, S., Das, G., & Wong, E. (2019b). Cost-optimal cloudlet placement frameworks over fiber-wireless access network for low-latency applications. *Journal of Network and Computer Applications*, *138*, 27–38. doi:10.1016/j.jnca.2019.04.014

Mukherjee, A., De, D., & Roy, D. G. (2019). A power and latency aware cloudlet selection strategy for multi-cloudlet environment. *IEEE Transactions on Cloud Computing*, *7*(1), 141–154. doi:10.1109/TCC.2016.2586061

Nayak, S. C., Parida, S., Tripathy, C., & Pattnaik, P. K. (2019). Dynamic backfilling algorithm to increase resource utilization in cloud computing. *International Journal of Information Technology and Web Engineering*, *14*(1), 1–26. doi:10.4018/IJITWE.2019010101

Pang, Z., Sun, L., Wang, Z., Tian, E., & Yang, S. (2015). A survey of cloudlet based mobile computing. In *2015 International Conference on Cloud Computing and Big Data (CCBD)* (pp. 268-275). IEEE. doi:10.1109/CCBD.2015.54

Raei, H., Ilkhani, E., & Nikooghadam, M. (2019). SeCARA: A security and cost-aware resource allocation method for mobile cloudlet systems. *Ad Hoc Networks*, *86*, 103–118. doi:10.1016/j.adhoc.2018.11.002

Rahimi, H., Picaud, Y., Costanzo, S., Madhusudan, G., Boissier, O., & Singh, K. D. (2020). *Design and Simulation of a Hybrid Architecture for Edge Computing in 5G and Beyond.* https://arxiv.org

Satyanarayanan, M., Bahl, P., Caceres, R., & Davies, N. (2009). The case for vm-based cloudlets in mobile computing. *IEEE Pervasive Computing*, *8*(4), 14–23. doi:10.1109/MPRV.2009.82

Shaukat, U., Ahmed, E., Anwar, Z., & Xia, F. (2016). Cloudlet deployment in local wireless networks: Motivation, architectures, applications, and open challenges. *Journal of Network and Computer Applications*, *62*(2), 18–40. doi:10.1016/j.jnca.2015.11.009

Shen, C., Xue, S., & Fu, S. C. (2019). ECPM: An energy-efficient cloudlet placement method in mobile cloud environment. *Journal on Wireless Communications and Networking*, *141*(1), 1–10. doi:10.1186/s13638-019-1455-8

Sun, X., & Ansari, N. (2019). Adaptive avatar handoff in the cloudlet network. *IEEE Transactions on Cloud Computing*, *7*(3), 664–676. doi:10.1109/TCC.2017.2701794

Tyng-Yeu, L., & You-Jie, L. (2017). A location-aware service deployment algorithm based on k-means for cloudlets. *Mobile Information Systems*, *2017*, 1–10. doi:10.1155/2017/8342859

Verbelen, T., Simoens, P., Turck, F. D., & Dhoedt, B. (2014). Adaptive deployment and configuration for mobile augmented reality in the cloudlet. *Journal of Network and Computer Applications*, *41*(1), 206–216. doi:10.1016/j.jnca.2013.12.002

Wang, Z., Gao, F., & Jin, X. (2020). Optimal deployment of cloudlets based on cost and latency in Internet of Things networks. *Wireless Networks*, *26*(8), 6077–6093. doi:10.1007/s11276-020-02418-9

Wang, Z., Zhao, D., Ni, M., Li, L., & Li, C. (2020). Collaborative Mobile Computation Offloading to Vehicle-based Cloudlets. *IEEE Transactions on Vehicular Technology*, *70*(1), 768–781. doi:10.1109/TVT.2020.3043296

Wei, H., Luo, H., & Sun, Y. (2020). Mobility-aware service caching in mobile edge computing for internet of things. *Sensors (Basel)*, *20*(3), 610–630. doi:10.3390/s20030610 PMID:31979135

Yang, S., Li, F., Shen, M., Chen, X., Fu, X., & Wang, Y. (2019). Cloudlet placement and task allocation in Mobile Edge Computing. *IEEE Internet of Things Journal*, *6*(3), 5853–5863. doi:10.1109/JIOT.2019.2907605

Yao, H., Bai, C., Xiong, M., Zeng, D., & Fu, Z. (2017). Heterogeneous cloudlet deployment and user-cloudlet association toward cost effective fog computing. *Concurrency and Computation*, *29*(16), 1–9. doi:10.1002/cpe.3975

Zhang, F., Ge, J., Li, Z., Li, C., Wong, C., Kong, L., Luo, B., & Chang, V. (2018). A load-aware resource allocation and task scheduling for the emerging cloudlet system. *Future Generation Computer Systems*, *87*(10), 438–456. doi:10.1016/j.future.2018.01.053

Zhang, Y., Jiao, L., Yan, J. Y., & Lin, X. J. (2019). Dynamic service placement for virtual reality group gaming on mobile edge cloudlets. *IEEE Journal on Selected Areas in Communications*, *37*(8), 1881–1897. doi:10.1109/JSAC.2019.2927071

Zhao, L., Sun, W., Shi, Y., & Liu, J. (2018). Optimal placement of cloudlets for access delay minimization in sdn-based internet of things networks. *IEEE Internet of Things Journal*, *5*(2), 1334–1344. doi:10.1109/JIOT.2018.2811808