Hindawi
Wireless Communications and Mobile Computing
Volume 2022, Article ID 4553317, 17 pages
https://doi.org/10.1155/2022/4553317

Research Article

SCOTT: Scheduling of Comprehensive Objectives for Tasks with Multitargets in Computing Networks

Guowei Zhang[1], **Zening Liu**[2], **Kunlun Wang**[3], **Xiaodong Zang**[1], **Yong Zuo**[4] and **Yang Yang**[5,6,7]

[1]*School of Cyber Science and Engineering, Qufu Normal University, Qufu 273165, China*
[2]*Purple Mountain Laboratories, Nanjing 211111, China*
[3]*School of Communication and Electronic Engineering, East China Normal University, Shanghai 200241, China*
[4]*College of Electronic Science and Technology, National University of Defense Technology, Hunan 410003, China*
[5]*Terminus Group, Beijing 100027, China*
[6]*Peng Cheng Laboratory, Shenzhen 518055, China*
[7]*Shenzhen Smart City Technology Development Group Co. Ltd., Shenzhen 518046, China*

Correspondence should be addressed to Guowei Zhang; zhanggw@qfnu.edu.cn

Received 16 March 2022; Accepted 29 August 2022; Published 27 September 2022

Academic Editor: Ihsan Ali

Local and customized services are realized with new type computing architecture by utilizing the spare resources distributed on the helper nodes (HNs) throughout the network. The heterogeneity of mobile edge and fog computing networks makes them natural to support multitarget tasks, and efficient task scheduling is always a fundamental and hot issue in multitask multihelper (MTMH) computing networks. Unlike most of the researches concentrating on the optimization of a single or limited service metrics, this article proposes a service framework for multitarget tasks, which is more universal for future 6G networks supporting customized services. The comprehensive quality of service (CQoS) is constructed to indicate the comprehensive objectives of the task nodes (TNs) with multiple targets. By formulating and transforming the CQoS maximal problem into two one-variable form subproblems, an algorithm named scheduling of comprehensive objectives for tasks with multitargets (SCOTT) is proposed. The SCOTT algorithm achieves the optimal offloading service solutions considering service metrics including delay, energy consumption, and economic cost. Extensive numerical simulations are carried out, which indicate that the proposed SCOTT algorithm can effectively achieve the optimal offloading solutions including node selection, task division, and transmission power for TNs with various service targets. Moreover, the universal applicability of the SCOTT algorithm is verified with case studies and numerical results.

1. Introduction

Benefitted from the development of the technologies including wireless communications and local processing, the world is entering the all-connect era, where the data exchange and computing transfer are ubiquitous [1]. Billions of terminal devices are connected to the network and construct the Internet of Things (IoT) systems, which leads to the exponential growth of the data traffic [2, 3]. Fully local processing cannot support applications of all scenes for the reason that local processing capabilities are limited [4]. In the well-developed traditional networks, tasks generated at the task nodes (TNs) that exceed the local processing capabilities can be transferred to the central cloud server [4, 5]. However, with the explosion of the mobile data generated by the emerging 5G and IoT applications, traditional centralized offloading architecture will bring a heavy burden to the link between the cloud server and the TNs. Besides, the increasing of the network scale and complexity makes the real-time processing and global optimization impractical and challenges the overall service quality [6, 7]. To cope with the problems emerged in the upcoming internet of everything, new network architectures that are more flexible and extensible need to be developed; thus, the massive and

diverse TNs can be efficiently supported, which are more universal for future 6G networks providing customized services.

Assisted by technologies including the network function virtualization (NFV) and the software defined network (SDN) [8], the concept of mobile computing has emerged. Mobile computing combines the cloud computing, edge computing [9], and fog computing [10] and provides computing for user services anywhere anytime. In a network supported by mobile computing technology, the networks resources including relaying, caching, and computing are subsided from the central server and extended to the whole network space [11, 12]. The resources are carried by the helper nodes (HNs) distributed throughout the network, which construct the multitask multihelper (MTMH) computing network together with the numerous TNs [13]. This architecture provides a rich collection of the ubiquitous network resources, and it has characteristics such as location awareness, widespread geographical distribution, huge number of nodes, and heterogeneity [14, 15]. For the tasks generated at anywhere of the network, their particular service targets can be satisfied by offloading all the data, or partially, to the proper HNs rather than the only choice, i.e., the central cloud server. Therefore, through jointly scheduling the HNs with various capabilities and TNs with various service targets, mobile computing can achieve better quality of service (QoS) for metrics such as delay, energy consumption, and security [16–18]. The heterogeneities and massiveness of the TNs and HNs make the optimal radio and computing resource management in computing networks challenging [19], which leads to extensive researches on the task scheduling in computing networks.

2. Related Works

Real-time performance is an ever-increasing requirement of the network services [20–23]. Better real-time performance can be achieved through the services provided by the nearby HNs, and this makes the task delay an important scheduling metric of the mobile computing services. Markov decision process approach is adopted to deal with computing task scheduling problem from the cloud to the mobile-edge computing (MEC) server, and delay optimal offloading solution is achieved under a power-constrained condition [20]. In [23], fog computing is integrated with vehicular networks, and a three-layer vehicular fog computing (VFC) mode is constructed to minimize the response time by leveraging moving and parked vehicles as fog nodes. A new hybrid offloading architecture for VFC is proposed in [24], and the node selection is optimized to reduce the offloading delay. In [25], the maximum delay among users in a mobile cloud computing system is minimized by randomization mapping method. In [26], the autonomy of the HNs is taken into consideration when pursing the delay-optimal offloading solution. Based on queueing theory, analytical model is introduced for service delay in [27], and the delay-minimized policy is provided when fog computing is introduced as a complement to cloud computing and an essential ingredient of the IoT. Considering the task scheduling problem for multitasks, low processing delay offloading solution for unsplittable tasks is achieved in [13], and the work is extended to splittable tasks in [28].

The decreasing of the distance between the user and server can dramatically reduce the transmission energy consumption [29, 30]. Meanwhile, the energy consumption is usually a sensitive metric in computing and IoT networks, for the reason that a large portion of the devices in IoT networks have a limited battery life [31–33]. This makes the energy efficiency an influential scheduling metric of the mobile computing services. The energy-efficient task offloading problem in mobile cloud computing networks is considered in [34], in which a distributed energy-efficient dynamic offloading and resource scheduling (eDors) algorithm is proposed. The eDors algorithm achieves the energy-efficient task offloading solution by simultaneously deciding the computation offloading selection, clock frequency control, and transmission power allocation. A task selection and scheduling scheme called CoESMS is introduced in [35], which minimizes the overall energy consumption and makespan through cooperative game theory models. In [33], a wireless powered MEC network architecture is proposed to support task offloading services, and energy-efficient offloading scheme is analyzed to support mobile devices with finite battery life. The authors of [36] proposed a scheduling strategy of the frequency division technique based on machine learning, which achieves good energy consumption minimization performance in mobile edge computation offloading.

In most cases, incentive is essential to get the services from HNs or the mobile service operators. From the view of the HNs, they want to make profit as much as possible based on their own capabilities. From the view of the TNs, they want to minimize their economic cost on condition that the services are satisfactory. Therefore, economic cost of is another important service metric when making scheduling decisions. Game theory is widely adopted in the researches of this area. In [37], a two-stage game in three-layer mobile crowd sensing (MCS) architecture is considered in edge computing networks, and a Markov decision process-(MDP-) based social model is built to achieve the maximal social welfare. A quality-aware traffic offloading (QATO) framework is proposed in [38], incentive schemes are adopted among neighbor nodes to achieve better service quality. Shen et al. [39] proposed an incentive framework for resource sensing based on the Stackelberg game, and the optimal solutions including sensing price and sensing frequency are derived. A trilateral game among service provider, end users, and edge resource owners is modeled in [40], and a two-stage dynamic game is used to evaluate the profit of each participant. In addition, service metrics such as fairness, security, and resilience are widely investigated in mobile computing networks [37, 41–43].

Tradeoff between different service metrics is also widely studied in this area. The performance indexes including task delay and energy consumption are abstracted to revenue and cost in the operation process of the fog-enabled computing network [44, 45], and game theories are adopted to achieve the balance of payments. In [46], a solution to the helper node location problem is provided, which provides support

for mobile users with limited battery while being able to process heavy workloads with low latency constraints. Yang et al. [47] proposed a low complexity algorithm that provides the maximal energy efficiency scheduling decisions under feasible modulation and time allocations. Tradeoff between energy consumption and task delay is achieved by Zhao et al. [48], in which the total energy consumption of multiple mobile devices is minimized subject to bounded-delay requirement. In [49], state-of-the-art studies for the joint wireless power transfer (WPT) and offloading in MEC are compared, and a taxonomy are formulated for the technologies that provide offloading service for smart devices while extending battery lifetime. The user mobility is considered in [50], and a lightweight mobility prediction and offloading (LiMPO) framework using artificial neural networks with less complexity is proposed, which achieves better performances in latency reduction, energy efficiency, and resource utilization.

Based on the above literature review, we find that most of the researches on the scheduling of computing services focus on the optimization of a single or very limited kinds of service metrics. However, the applications in future 6G and IoT networks always have various service targets. In addition, mobile computing is always in heterogeneous and MTMH style, which is appropriate for the processing of multitarget tasks. The tasks that are sensitive to different metrics are scheduled together, which have different tendencies of node selection and offloading strategy. To cope with this, it is of great necessary to develop universal task scheduling scheme in computing networks. Thus, the heterogeneous resources distribute on the HNs can be integrated to provide customized services for the comprehensive service objectives of the TNs.

Therefore, the main contributions of this paper are summarized as follows:

(1) We propose a general service model for multitarget tasks in MTMH computing networks. Heterogeneous service capabilities of the HNs and the various service targets of the TNs are collected at the scheduler. The comprehensive objective of the service is formulated as the comprehensive QoS (CQoS) by weighting the absolute service metrics with service target factors, and scheduling scheme is made aiming to maximize the CQoS

(2) Considering the service metrics including task delay, energy consumption, and economic cost, the CQoS maximal problem for the offloading service with three targets is formulated. By transforming the original problem into two one-variable form subproblems, we develop an algorithm named scheduling of comprehensive objectives for tasks with multitargets (SCOTT), which provide the optimal offloading solution including node selection, task division, and transmission power. Case studies are conducted out to further prove the practicability of our proved offloading scheme

(3) Extensive simulations in a computing network are carried out to investigate the performance of our proposed scheduling algorithm. Numerical results show that the SCOTT algorithm can effectively obtain the CQoS maximal offloading solution based on multiple available HNs and provide the optimal offloading services for TNs with various targets in different network scenarios

The rest of this paper is organized as follows. The general service model for multitarget tasks is introduced in Section 3. In Section 4, we formulate the CQoS and the corresponding optimization problem of the offloading service concerning metrics including delay, energy consumption, and economic cost. In Section 5, the CQoS maximal problem for 3-target tasks is solved, and the SCOTT algorithm is proposed, which provides the optimal offloading solution including node selection, task division, and transmission power. Case studies for the proposed SCOTT algorithm are carried out in Section 6. The numerical estimations are provided in Section 7. Section 8 concludes this paper.

3. Service Model for Multitarget Tasks

In this section, a general MTMH computing network supporting tasks with N targets is introduced. The service capabilities, service objectives, and service scheme are intergraded to represent the comprehensive satisfactory level of the offloading service, which can guide the direction to achieve the optimal service solutions.

3.1. Task Scheduling in MTMH Computing Networks. As shown in Figure 1, we consider an MTMH mobile computing network consisting of multiple TNs and HNs, which have various service targets and service capabilities. A task scheduler in this network collects these service capabilities from the HNs and the service requests from the TNs and provides the service scheme. The scheduler may be located at the cloud server or a specific HN. In this mobile computing network, the task generated at the TN with size l can be offloaded to the nearby HNs to achieve better task processing quality. The task processing targets of different TNs are always diverse, and the task processing target of a same user can be time-varying. For the service provided by a certain HN, the relationship between the service capabilities and the service targets can be revealed by the goodness of fit between the HN and TN characteristics, which is illustrated in Figure 1. This goodness of fit can reflect the comprehensive satisfactory level of the service, and it depends on the following three elements:

(1) The service capabilities of the M HNs, which can be represented by $\Lambda = \{\Lambda_1, \Lambda_2, \cdots \Lambda_M\}$. The service capabilities of HN i, i.e., Λ_i, may consist of the service rate, service energy consumption, service price, and any other elements related to the service process

(2) The comprehensive service objective of the TN with multitargets, which can be represented by $K = \{K_1, K_2, \cdots, K_N\}$. The parameter N is the number of the service metrics such as delay and energy

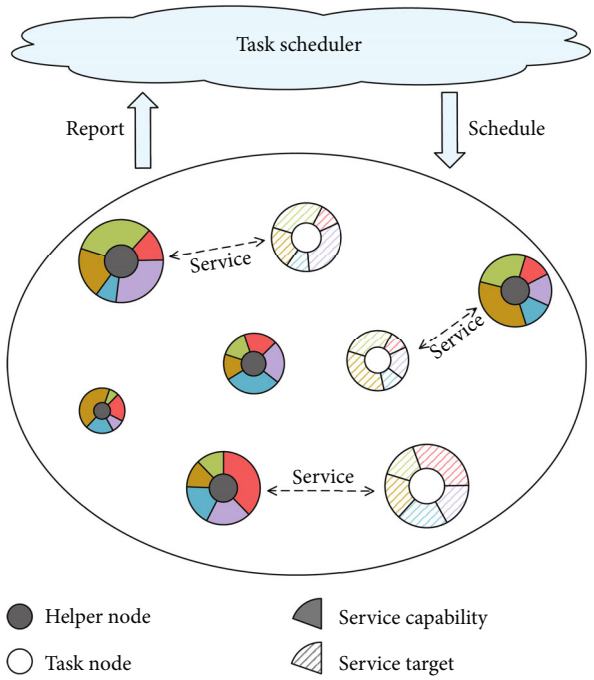

Figure 1: Service framework for multitarget tasks in an MTMH computing network. The service capabilities of the HNs and the service requests of the TNs are collected at the task scheduler, which makes the service scheme for the tasks with multitarget.

consumption. The larger the factor K_n is, the more sensitive the TN is to the corresponding service metric

(3) The service scheme provided by the scheduler, which can be represented by $O = \{i, l_i, p_i\}$. The parameters i, l_i, and p_i are the index of the selected HN, the size of the subtask offloaded to the selected HN, and the transmission power of the TN, respectively

3.2. Comprehensive QoS. For a specific task, the service scheme O is provided based on the collected service capability Λ. Then, the absolute service metrics such as service delay and cost can be achieved, which is denoted by $S = \{S_1, S_2, \cdots, S_N\}$. Therefore, we have

$$S = \xi(\Lambda, O), \tag{1}$$

where ξ is the map function between the service capabilities/scheme and the absolute service metrics.

The task has different sensitivities to different service metrics, which are quantized by $K_n, n = 1, 2, \cdots, N$. In order to estimate the service quality in a general way, the absolute service metric S_n is weighted by the corresponding service objective factor K_n. Then, the summation of the weighted service metrics can be calculated by the scalar product of K and S, i.e.,

$$Q = K \bullet S = K \bullet \xi(\Lambda, O). \tag{2}$$

Based on the service scheme O and the service capabilities Λ, we define the weighted summation Q as the CQoS of the service provided for task with service objective K, and this integrated metric Q reveals the goodness of fit between the service provider and service requester.

Taking a look at the expression of CQoS in (2), we find that the only variable is the service scheme S. Therefore, the general optimization problem that achieves the CQoS maximal service scheme O^* can be formulated as

$$\begin{aligned} \mathbf{P}: \max_{O} \; & Q = K \bullet \xi(\Lambda, O), \\ \text{s.t.} \; & i \in \mathcal{F}, \\ & 0 \le l_i \le l, \\ & 0 \le p_i \le p_{\max}, \end{aligned} \tag{3}$$

where $\mathcal{F}$ is the set of the available HNs in this computing network.

For any TN with various service objectives in a heterogeneous computing network, the optimization problem **P** provides the direction to search the CQoS maximal service scheme, no matter what the service capabilities Λ and the service objectives K are. This demonstrates the universal applicability of this service framework.

4. Offloading Service for 3-Target Tasks

Based on the proposed service framework, the rest of the paper concentrates on the offloading service for 3-target tasks. The service metrics including delay, energy consumption, and economic cost are investigated.

4.1. Metric Formulation. We use $K = \{K_d, K_e, K_c\}$ to denote the comprehensive service objective of a TN, in which K_d, K_e, and K_c with nonnegative values are the delay factor, energy consumption factor, and the cost factor, respectively. The higher a factor in K is, the more sensitive the task is to the corresponding service metric. In particular, the task with $K_d \neq 0$ and $K_e = K_c = 0$ is completely delay-sensitive, and the delay-minimized offloading scheme achieves the highest service quality for this task.

The service capabilities of an available HN, say HN i, is specified as $\Lambda_i = \{f_i, \eta_i, \theta_i, \pi_i\}$; the explanations of the parameters are summarized in Table 1. The task offloading scheme provided by the task scheduler, i.e., $O = \{i, l_i, p_i\}$, includes the HN selection, task division (l_T, l_i), and the task transmission power p_i. In other words, the scheduler needs to select a proper HN, determine the offload data size, and provide the optimal transmission power from TN to the selected HN.

Next, we formulate the delay D_i, energy consumption E_i, and economic cost C_i based on Λ_i and O. Thus, we can get the absolute service metrics as $S = \{D_i, E_i, C_i\}$.

4.1.1. Task Offloading Delay. The overall task offloading delay when HN i is selected includes two parts, i.e., the delay of the local subtask with l_T bits and the delay of the offloaded

Table 1: Summary of key notations.

Nota.	Unit	Description
l	Bit	Overall task size of the TN.
l_T	Bit	Subtask size processed locally at the TN.
l_i	Bit	Subtask size offloaded to HN i.
η_T	Cycle/bit	CPU cycles for processing 1 bit data at the TN.
η_i	Cycle/bit	CPU cycles for processing 1 bit data at HN i.
f_T	Cycle/s	CPU frequency of the TN.
f_i	Cycle/s	CPU frequency of HN i.
θ_T	J/cycle	Energy consumption per CPU cycle of the TN.
θ_i	J/cycle	Energy consumption per CPU cycle of HN i.
π_i	\$/bit	The service price of HN i.
W	Hz	Spectrum bandwidth for task offloading.
p_i	J/s	Transmission power of the TN to HN i.
p_{max}	J/s	Upper bound of the transmission power of the TN.
γ_i	–	Path loss factor between the TN and HN i.
β_i	–	Shadowing factor between the TN and HN i.
I_i	J/s	Interference power between the TN and HN i.
N_0	J/(s·Hz)	Noise power spectral density.
K_d	/s	Delay factor of the comprehensive service objective.
K_e	/J	Energy consumption factor of the comprehensive service objective.
K_c	/\$	Cost factor of the comprehensive service objective.
K	–	Comprehensive service objective.
D_i	s	Task offloading delay through HN i.
E_i	J	Task offloading energy consumption through HN i.
C_i	\$	Task offloading economic cost through HN i.
Q_i	–	Local CQoS of the offloading service through HN i.

subtask with l_i bits, which are denoted by D_{Ti} and D_{Oi}, respectively. In most applications, the overall delay of the task offloading service is decided by the maximum processing time of the subtasks; thus, we have

$$D_i = \max\left(D_{Ti}, D_{Oi}\right). \tag{4}$$

Following the model in our previous research [41], the time of processing 1 bit data locally is η_T/f_T, in which f_T is the CPU frequency of the TN, and η_T is the CPU cycles for processing 1 bit data at the TN. Thus, the local delay D_{Ti} can be expressed as

$$D_{Ti} = l_T \frac{\eta_T}{f_T}. \tag{5}$$

Compared with the local delay, the offloading delay D_{Oi} involves the processing delay in similar form to D_{Ti}, and a transmission delay in addition, i.e.,

$$D_{Oi} = \frac{l_i \eta_i}{f_i} + \frac{l_i}{WB_i} = l_i\left(\frac{1}{WB_i} + \frac{\eta_i}{f_i}\right), \tag{6}$$

where f_i is the CPU frequency of HN i, η_i is the CPU cycles for processing 1 bit data at HN i, W is the spectrum bandwidth allocated to the offloading service, and B_i is the spectral efficiency of the wireless link from the TN to HN i. Given the terminal transmission power p_i, B_i is obtained through the Shannon capacity as

$$B_i = \log_2\left(1 + \frac{p_i \gamma_i \beta_i}{I_i + WN_0}\right). \tag{7}$$

In the expression of B_i, γ_i and β_i are the path loss and shadowing factors of this wireless link. I_i and N_0 are the interference power and the noise power spectral density, respectively.

4.1.2. Task Offloading Energy Consumption. The overall offloading energy consumption E_i includes the computing energy consumption and the transmission energy consumption. In this research, we use θ_T and θ_i to represent the energy consumption per CPU cycle of the TN and HN i. Then, the computing energy consumptions per bit data for the TN and HN i are represented by $\eta_T\theta_T$ and $\eta_i\theta_i$, respectively. Taking the transmission energy consumption with transmission power p_i into consideration, the overall energy consumption is formulated as

$$E_i = E_{T,i} + E_{Oi} = l_T\eta_T\theta_T + \frac{l_i p_i}{WB_i} + l_i\eta_i\theta_i, \tag{8}$$

where $E_{T,i} = l_T\eta_T\theta_T + (l_i p_i/WB_i)$ is the energy consumptions of the TN, and $E_{O,i} = l_i\eta_i\theta_i$ is the offloading energy consumption when HN i is selected.

4.1.3. Task Offloading Cost. Remunerations are requisite in most computing network applications, regardless of the HNs are individual devices with spare resources or specially deployed by operators. The economic cost of the task offloading service is usually proportional to the offloading data size. We use a parameter π_i to denote the remuneration of HN i when one bit data is offloaded to it. Therefore, the economic cost of the offloading service with offloading task size l_i is

$$C_i = l_i\pi_i. \tag{9}$$

Based on the metrics of the task offloading service including D_i, E_i, and C_i, the CQoS and the optimization problem need to be specified.

4.2. CQoS and Optimization Problem. The satisfaction degree of the TN for the service provided by the computing network depends on both the service target of the task itself

and the service capabilities of the HNs. For a TN with comprehensive service objective $K = \{K_d, K_e, K_c\}$, the weighted offloading service metrics are K_dD_i, K_eE_i, and K_cC_i. Then, we construct the CQoS of the offloading service provided by HN i as

$$Q_i = \frac{1}{K_dD_i + K_eE_i + K_cC_i}, \tag{10}$$

which indicates that task offloading schemes with low delay, low energy consumption, and low economic cost can achieve high comprehensive service qualities, and the impact of specific service targets are weighted by the corresponding factors. This is also the reason why there is a reciprocal in (10).

Given a selected HN, the CQoS provided above is taken as the utility of the provided service, and it is directly decided by the subtask size l_i and the TN transmission power p_i. Therefore, we propose the following optimization problem.

$$\begin{aligned} \mathbf{P0}: \max_{l_i,p_i}\ & Q_i \\ \text{s.t.}\ & 0 \le l_i \le l \\ & 0 \le p_i \le p_{\max}, \end{aligned} \tag{11}$$

where $p_{\max}$ is the upper bound of TN transmission power. For each available HN, we need to solve the corresponding optimization problem to find the local optimal offloading solution (l_i^*, p_i^*) and the corresponding local maximal CQoS Q_i^*, $i \in \mathcal{F}$. In this way, the global optimal offloading solution $O^* = (i^*, l_{i^*}^*, p_{i^*}^*)$ and global maximal CQoS Q^* can be obtained by selecting the HN with the highest local maximal CQoS. This scheme is applicable to TNs with differentiated service targets.

5. SCOTT Algorithm

In this section, we propose the SCOTT algorithm for the scheduling of the 3-target tasks, which solves the CQoS optimization problem by transformed into two one-variable form subproblems. Thus, the global optimal offloading solution $O^* = (i^*, l_{i^*}^*, p_{i^*}^*)$ and global maximal CQoS Q^* are obtained.

5.1. Problem Transformation. For the original optimization problem **P0**, the maximization of the local CQoS Q_i is equivalent to the minimization of the denominator in (10). Therefore, we can transform **P0** into **P1** as

$$\begin{aligned} \mathbf{P1}: \min_{l_i,p_i}\ & Q_{i1} = K_dD_i + K_eE_i + K_cC_i \\ \text{s.t.}\ & 0 \le l_i \le l \\ & 0 \le p_i \le p_{\max}, \end{aligned} \tag{12}$$

in which $Q_{i1} = Q_i^{-1}$.

As defined in (4), the overall delay D_i of the task offloading service provided by HN i is decided by the larger subtask delay. Then, we have the following proposition.

Proposition 1. *When the CQoS Q_i is maximized under the condition that $D_i = \max(D_{T,i}, D_{Oi})$, the offloading delay D_{Oi} is no less than the local delay $D_{T,i}$.*

Proof. Please refer to Appendix A. □

According to Proposition 1, the delay target in **P0** and **P1** can be represented as K_dD_{Oi}, and the subtask size l_i should satisfies

$$\begin{aligned} D_{Ti} \le D_{Oi} &\Longrightarrow (l - l_i)\frac{\eta_T}{f_T} \\ &\le l_i\left(\frac{1}{WB_i} + \frac{\eta_i}{f_i}\right) \Longrightarrow l\frac{\eta_T}{f_T}\left(\frac{\eta_T}{f_T} + \frac{1}{WB_i} + \frac{\eta_i}{f_i}\right)^{-1} \le l_i < l. \end{aligned} \tag{13}$$

Therefore, **P1** can be transformed into

$$\begin{aligned} \mathbf{P2}: \min_{l_i,p_i}\ & Q_{i2} = K_dD_{Oi} + K_eE_i + K_cC_i \\ &= K_dl_i\left(\frac{1}{WB_i} + \frac{\eta_i}{f_i}\right) + K_cl_i\pi_i \\ &\quad + K_e\left[l\eta_T\theta_T + l_i\left(\frac{p_i}{WB_i} + \eta_i\theta_i - \eta_T\theta_T\right)\right] \\ \text{s.t.}\ & l\frac{\eta_T}{f_T}\left(\frac{\eta_T}{f_T} + \frac{1}{WB_i} + \frac{\eta_i}{f_i}\right)^{-1} \le l_i \le l \\ & 0 \le p_i \le p_{\max}. \end{aligned} \tag{14}$$

Remark 2. The conclusions in Proposition 1 and **P2** can be explained as follows. For a certain HN and a subtask size l_i, the increase of the transmission power p_i cannot continually decrease the overall task offloading delay, for the reason that the local processing capability of the TN is limited. Besides, a larger transmission power will undoubtedly lead to a higher energy consumption and thus a lower CQoS. The correlation between l_i and p_i in the CQoS maximization problem is revealed by (13).

It is easy to know that problem **P2** is not convex. Then, we will further transform **P2** into one-variable form to find the optimal offloading solution. For the optimization objective Q_{i2}, the first derivative with respect to l_i is

$$\frac{\partial Q_{i2}}{\partial l_i} = K_d\left(\frac{1}{WB_i} + \frac{\eta_i}{f_i}\right) + K_e\left(\frac{p_i}{WB_i} + \eta_i\theta_i - \eta_T\theta_T\right) + K_c\pi_i. \tag{15}$$

The derivative $\partial Q_{i2}/\partial l_i$ is uncorrelated with l_i. At the mean time, given the comprehensive service target $K = \{K_d, K_e, K_c\}$ of the TN and the service capabilities of the HN i, $\partial Q_{i2}/\partial l_i$ is directly decided by the TN transmission power p_i. As a result, we can divide the value range of p_i in **P2**, i.e., $[0, p_{\max}]$, into two parts, in which $\partial Q_{i2}/\partial l_i$ is nonnegative and negative, respectively. As shown below, these two value ranges are denoted as $\mathcal{R}_1$ and $\mathcal{R}_2$, respectively.

$$\mathcal{R}_1 = [0, p_{\max}] \cap \left\{p_i \middle| \frac{\partial Q_{i2}}{\partial l_i} \ge 0\right\}, \tag{16}$$

$$\mathscr{R}_2 = [0, p_{\max}] \cap \left\{ p_i \middle| \frac{\partial Q_{i2}}{\partial l_i} < 0 \right\}. \tag{17}$$

When the subtask size l_i increases, the minimization goal Q_{i2} in **P2** is severally monotone increasing in $\mathscr{R}_1$ and monotone decreasing in $\mathscr{R}_2$. Consequently, the optimal value of l_i that minimizes Q_{i2} in $\mathscr{R}_1$ and $\mathscr{R}_2$ should be $l(\eta_T/f_T)((\eta_T/f_T) + (1/WB_i) + (\eta_i/f_i))^{-1}$ and l, respectively.

Based on the above analysis, we can transform the optimization problem **P2** into two subproblems as follows:

$$\mathbf{P2-1}: \tag{18}$$

$$\min_{p_i} Q_{i21} = (K_d D_{Oi} + K_e E_i + K_c C_i)\big|_{l_i = l(\eta_T/f_T)((\eta_T/f_T)+(1/WB_i)+(\eta_i/f_i))^{-1}} \tag{19}$$

$$\text{s.t.} p_i \in \mathscr{R}_1. \tag{20}$$

$$\mathbf{P2-2}: \tag{21}$$

$$\min_{p_i} Q_{i22} = (K_d D_{Oi} + K_e E_i + K_c C_i)\big|_{l_i = l} \tag{22}$$

$$\text{s.t.} p_i \in \mathscr{R}_2. \tag{23}$$

Obviously, **P2 − 1** and **P2 − 2** are both one-variable form optimization problems. We can solve the original problem **P0** by firstly finding the optimal solutions of these two subproblems severally, then achieving optimal CQoS Q_i^* based on Q_{i21}^* and Q_{i22}^*. The HN with the highest Q_i^* will be selected as the helper node, and the corresponding task division and transmission power will also be achieved.

5.2. Subproblem Solving. To solve the two subproblems obtained above, the two corresponding value ranges, i.e., $\mathscr{R}_1$ and $\mathscr{R}_2$, need to be determined first.

The value ranges of **P2 − 1** and **P2 − 2** are determined by the sign of the gradient $\partial Q_{i2}/\partial l_i$ (16). Therefore, the following proposition is provided.

Proposition 3. *The gradient $\partial Q_{i2}/\partial l_i$ is monotone decreasing when $p_i \in [0, \widehat{p}_i]$ and monotone increasing when $p_i \in (\widehat{p}_i, \infty)$, in which $\widehat{p}_i$ is the only positive solution of the following equation.*

$$K_e B_i - \frac{\gamma_i \beta_i / (I_i + WN_0)}{(1 + (p_i \gamma_i \beta_i / (I_i + WN_0)) I_i + WN_0)) \ln 2} (K_d + K_e p_i) = 0. \tag{24}$$

Proof. Please refer to Appendix B. □

According to the conclusions in Proposition 3, $\partial Q_{i2}/\partial l_i$ achieves its minimum value when $p_i = \widehat{p}_i$. If $\partial Q_{i2}/\partial l_i|_{\widehat{p}_i} \geq 0$, $\partial Q_{i2}/\partial l_i$ is always nonnegative when $p_i \in [0, \infty)$. Otherwise, $\partial Q_{i2}/\partial l_i$ has two zero points in $(0, \infty)$, which are severally denoted by $`p_i$ and $p'_i$ ($`p_i < p'_i$). Then, the sign of $\partial Q_{i2}/\partial l_i$ can be summarized as:

When $\partial Q_{i2}/\partial l_i|_{\widehat{p}_i} \geq 0$, $\partial Q_{i2}/\partial l_i \geq 0$ in $[0, \infty)$.

When $\partial Q_{i2}/\partial l_i|_{\widehat{p}_i} < 0$, $\partial Q_{i2}/\partial l_i \geq 0$ in $[0, `p_i]$ and $[p'_i, \infty)$ and $\partial Q_{i2}/\partial l_i < 0$ in $(`p_i, p'_i)$.

Taking the value range of p_i in the original optimization problem **P0**, i.e., $[0, p_{\max}]$, into consideration (16), a subalgorithm is introduced in Algorithm 1 to achieve $\mathscr{R}_1$ and $\mathscr{R}_2$.

We need to solve the two subproblems based on the values ranges achieved by Algorithm 1. First, the following proposition are provided for the optimal solutions of **P2 − 1**.

Proposition 4. *The optimal transmission power p_{i1}^* of the subproblem **P2 − 1** and the corresponding optimal subtask size l_{i1}^* are provided as follows:*

$$p_{i1}^* = \begin{cases} `p_i, & \mathscr{R}_1 \neq [0, p_{\max}], \\ \underset{p_{i1} \in \mathscr{P}_{i1}}{argmin}\, Q_{i21}, & \mathscr{R}_1 = [0, p_{\max}], \end{cases} \tag{25}$$

$$l_{i1}^* = l \frac{\eta_T}{f_T} \left(\frac{\eta_T}{f_T} + \frac{1}{WB_i} + \frac{\eta_i}{f_i} \right)^{-1} \Bigg|_{p_i = p_{i1}^*}, \tag{26}$$

In (25), $\mathscr{P}_{i1} = \{0, p_{\max}, \check{p}_i\}$ when the following three conditions are satisfied. Otherwise, $\mathscr{P}_{i1} = \{0, p_{\max}\}$.

$$\begin{cases} \dfrac{\alpha_2 \ln 2}{\alpha_3} < 1, \\ f(1) < 0, \\ f\left(1 + \dfrac{p \max \gamma_i \beta_i}{I_i + WN_0}\right) > 0. \end{cases} \tag{27}$$

The function $f(x)$ is

$$f(x) = \alpha_1 + \frac{\alpha_2}{x} + \alpha_3 \log_2(x), \tag{28}$$

in which the three parameters, i.e., α_1, α_2, and α_3, are

$$\begin{cases} \alpha_1 = -K_e \left(\dfrac{\eta_T}{f_T} + \dfrac{\eta_i}{f_i} \right) \dfrac{W}{\ln 2} + K_e, \\ \alpha_2 = \left\{ K_e \left(\dfrac{\eta_T}{f_T} + \dfrac{\eta_i}{f_i} \right) - \left[K_d \dfrac{\eta_T}{f_T} - K_e (\eta_i \theta_i - \eta_T \theta_T) - K_c \pi_i \right] \dfrac{\gamma_i \beta_i}{I_i + WN_0} \right\} \dfrac{W}{\ln 2}, \\ \alpha_3 = K_e \left(\dfrac{\eta_T}{f_T} + \dfrac{\eta_i}{f_i} \right) W > 0. \end{cases} \tag{29}$$

1: Initialize $\mathscr{R}_1 = \varnothing$, $\mathscr{R}_2 = \varnothing$, $\widehat{p}_i = \text{`}p_i = p'_i = 0$;
2: According to the system parameters, calculate the value of $\widehat{p}_i$ from
$K_e B_i - (\gamma_i \beta_i / (I_i + WN_0)) / (1 + (p_i \gamma_i \beta_i / (I_i + WN_0))) \ln 2)(K_d + K_e p_i) = 0$
with bisection method;
3: **if** $\partial Q_{i2} / \partial l_i|_{\widehat{p}_i} \geq 0$ **then**
4: $\mathscr{R}_1 = [0, p_{\max}]$, $\mathscr{R}_2 = \varnothing$;
5: **else**
6: Calculate the value of $\text{`}p_i$ and p'_i from
$K_d((1/WB_i) + (\eta_i / f_i)) + K_e((p_i / WB_i) + \eta_i \theta_i - \eta_T \theta_T) + K_c \pi_i = 0$
with bisection method;
7: **if** $p_{\max} \leq \text{`}p_i$ **then**
8: $\mathscr{R}_1 = [0, p_{\max}]$, $\mathscr{R}_2 = \varnothing$;
9: **end if**
10: **if** $\text{`}p_i < p_{\max} < p'_i$ **then**
11: $\mathscr{R}_1 = [0, \text{`}p_i]$, $\mathscr{R}_2 = [p'_i, p_{\max}]$;
12: **end if**
13: **if** $p_{\max} \geq p'_i$ **then**
14: $\mathscr{R}_1 = [0, \text{`}p_i] \cap [p'_i, p_{\max}]$, $\mathscr{R}_2 = [\text{`}p_i, p'_i]$;
15: **end if**
16: **end if**
17: **return** $\mathscr{R}_1, \mathscr{R}_2, \widehat{p}_i, \text{`}p_i, p'_i$;

Algorithm 1: Achieving value ranges.

When $\mathscr{P}_{i1}$ is $\{0, p_{\max}, \check{p}_i\}$, $\check{p}_i$ is the only solution of $f(1 + (p_i \gamma_i \beta_i / (I_i + WN_0))) = 0$ valued in $[0, p_{\max}]$.

Proof. Please refer to Appendix C. □

When the value range of **P2 − 2** is not empty, the following proposition provides the optimal offloading solution for this subproblem, as well as the relationship between Q^*_{i21} and Q^*_{i22}.

Proposition 5. *When $\mathscr{R}_2 \neq \varnothing$, the optimal transmission power p^*_{i2} of the subproblem **P2 − 2** and the corresponding optimal subtask size l^*_{i2} are provided as follows.*

$$p^*_{i2} = \begin{cases} p_{\max}, & p_{\max} \leq \widehat{p}_i, \\ \widehat{p}_i, & p_{\max} > \widehat{p}_i, \end{cases} \tag{30}$$

$$l^*_{i2} = l. \tag{31}$$

In addition, there is $Q^*_{i21} \geq Q^*_{i22}$ when $\mathscr{R}_2 \neq \varnothing$.

Proof. Please refer to Appendix D. □

Remark 6. The conclusions in the above two propositions can be intuitively explained as follows. If the value range of **P2 − 2** is not empty, the offloading scheme that achieves the highest CQoS has $l^*_o = l$. This means that in the value range $\mathscr{R}_2$, it is better to offload all the entire task to the HN because of the high cost performance of the offloading service. According to the expression of $\partial Q_{i2} / \partial l_i$ in (15), this high cost performance may come from the low processing energy consumption, the low service price, or the high energy consumption factor of the task.

Based on the value ranges achieved by Algorithm 1, Proposition 4 and 5 provide the optimal offloading solution (l^*_{i1}, p^*_{i1}), (l^*_{i2}, p^*_{i2}) and minimized utilities Q^*_{i21}, Q^*_{i22} of the two subproblems. The local optimal offloading solution (l^*_i, p^*_i) and the corresponding local maximal CQoS Q^*_i are given by

$$Q^*_i = \begin{cases} \dfrac{1}{Q^*_{i21}}, & \mathscr{R}_2 = \varnothing, \\ \dfrac{1}{Q^*_{i22}}, & \mathscr{R}_2 \neq \varnothing, \end{cases} \tag{32}$$

$$(l^*_i, p^*_i) = \begin{cases} (l^*_{i1}, p^*_{i1}), & \mathscr{R}_2 = \varnothing, \\ (l^*_{i2}, p^*_{i2}), & \mathscr{R}_2 \neq \varnothing. \end{cases}$$

This process is introduced in Algorithm 2.

5.3. Optimal Offloading Solution. Given a certain HN, the local optimal offloading solution and the local maximal CQoS are provided by Algorithms 1 and 2. In order to achieve the global CQoS maximal offloading solution $O^* = (i^*, l^*_{i^*}, p^*_{i^*})$ and the corresponding Q^*, we propose the SCOTT algorithm in Algorithm 3, in which the HN with the highest local maximal CQoS are selected.

6. Case Study for SCOTT Algorithm

Now, we investigate the task offloading services of several special cases, which are compared with the existing researches. Thus, the universal applicability of our proposed SCOTT task offloading scheme is further proved.

1: **Input** $\mathscr{R}_1$, $\mathscr{R}_2$, $\widehat{p}_i$, $'p_i$, p'_i;
2: **if** $\mathscr{R}_2 = \varnothing$ **then**
3: Calculate the optimal offloading solutions of **P2 – 1**, i.e., l^*_{i1}, p^*_{i1}, with (25) and (26);
4: Calculate the minimized utilities of **P2 – 1** by
$Q^*_{i21} = (K_d D_{Oi} + K_e E_i + K_c C_i)|_{l_i = l^*_{i1}, p_i = p^*_{i1}}$;
5: Set $l^*_i = l^*_{i1}$, $p^*_i = p^*_{i1}$, $Q^*_i = (Q^*_{i21})^{-1}$;
6: **else**
7: Calculate the optimal offloading solutions of **P2 – 2**, i.e., l^*_{i2}, p^*_{i2}, with (30) and (31);
8: Calculate the minimized utilities of **P2 – 2** by
$Q^*_{i22} = (K_d D_{Oi} + K_e E_i + K_c C_i)|_{l_i = l^*_{i2}, p_i = p^*_{i2}}$;
9: Set $l^*_i = l^*_{i2}$, $p^*_i = p^*_{i2}$, $Q^*_i = (Q^*_{i22})^{-1}$;
10: **end if**
11: **return** (l^*_i, p^*_i), Q^*_i;

Algorithm 2: Local CQoS maximal algorithm.

1: Initialize $\mathscr{F}$, $\mathscr{Q}$, $\mathscr{L}$, $\mathscr{P}$ as the sets of available HNs, local maximal CQoS, local optimal task division, and the local optimal TN transmission power.
2: **while** A task is generated **do**
3: Acquire the TN's service target factor $K = \{K_d, K_e, K_c\}$;
4: **for** each HN $i \in \mathscr{F}$ **do**
5: Call Algorithm 1;
6: Call Algorithm 2;
7: Update $\mathscr{Q} = \mathscr{Q} \cup Q^*_i$;
8: Update $\mathscr{L} = \mathscr{L} \cup l^*_i$;
9: Update $\mathscr{P} = \mathscr{P} \cup p^*_i$;
10: **end for**
11: Get the optimal HN: $i^* = \underset{i \in \mathscr{F}}{\operatorname{argmax}} \mathscr{Q}$;
12: Get the optimal offloaded task size: $l^*_{i^*} = \underset{l^*_i \in \mathscr{L}}{\operatorname{argmax}} \mathscr{Q}$;
13: Get the optimal offloading power: $p^*_{i^*} = \underset{p^*_i \in \mathscr{P}}{\operatorname{argmax}} \mathscr{Q}$;
14: Get the maximal CQoS: $Q^* = \max(\mathscr{Q})$;
15: **return** $O^* = (i^*, l^*_{i^*}, p^*_{i^*})$, Q^*;
16: **end while**

Algorithm 3: SCOTT algorithm.

6.1. Local Processing. The aim of calling for offloading service is to achieve a higher CQoS than local processing. If the cost performance of the offloading service is too low, the TN tends to abandon the offloading service and process the task locally. This happens when the processing efficiency of the corresponding HN is too low, or the economic cost is too high. The lower bound of the CQoS Q is

$$\underline{Q} = \frac{1}{K_d D_i + K_e E_i + K_c C_i} = \frac{1}{K_d l(\eta_T / f_T) + K_e l \eta_T \theta_T}, \quad (33)$$

which corresponds to the offloading solution $(l_i = 0, p_i = 0)$, i.e., local processing.

Substituting this solution into (10), we get $Q_i = K_d l(\eta_T / f_T) + K_e l \eta_T \theta_T = 1/\underline{Q}$. The SCOTT algorithm minimizes Q_i. This guarantees that the offloading solution obtained by our proposed SCOTT algorithms can always achieve the Q^* that is no less than $\underline{Q}$. Therefore, the SCOTT algorithm is applicable to local processing.

6.2. Delay-Sensitive Tasks. If the comprehensive service objective of a TN is $K = \{K_d > 0, K_e = 0, K_c = 0\}$, this task is a delay-sensitive task. This kind of task is not sensitive to the energy consumption or economic cost of the offloading service and focuses on the delay performance.

For this kind of tasks, the optimization goals of **P2** and **P2 – 1** become $K_d l_i((1/WB_i) + (\eta_i / f_i))$ and $K_d(l(\eta_T / f_T)\eta_T / f_T((1/WB_i)1/WB_i + (\eta_i / f_i)\eta_i / f_i))/((\eta_T / f_T) + (1/WB_i) + (\eta_i / f_i))$, respectively. Besides, the value ranges $\mathscr{R}_1$ and $\mathscr{R}_2$ are obviously $[0, p_{\max}]$ and $\varnothing$. For delay-sensitive tasks, the SCOTT algorithm achieves the local optimal offloading solution based on the conclusions in Proposition 4. Therefore, the SCOTT algorithm is applicable to delay sensitive tasks. This case corresponds to the problem solved in [26], in which task delay is the minimization goal.

6.3. Economic Cost-Sensitive Tasks. If the comprehensive service objective of a TN is $K = \{K_d = 0, K_e = 0, K_c > 0\}$, this task is a economic cost-sensitive task. This kind of task is not sensitive to the processing delay or energy consumption of the offloading service and focuses on the economic cost performance.

For this kind of tasks, local processing is the optimal solution, which can achieve a infinitely high CQoS. The optimization goal of **P2** becomes $K_c l_i \pi_i$, which is obviously proportional to the offloading task size l_i. Besides, the value ranges $\mathscr{R}_1$ and $\mathscr{R}_2$ are obviously $[0, p_{\max}]$ and $\varnothing$. Based on Proposition 4, the conditions in (27) are not fully satisfied, and $Q_i 21$ is a monotone increasing function of p_i. Therefore, the SCOTT algorithms will provide $p_i^* = 0$ and $l_i^* = 0$ in this case. Therefore, the SCOTT algorithm is applicable to economic cost sensitive tasks. This case can be applied to the case in [39], and the resource sensing frequency corresponds to the offloading task size in this paper.

6.4. Economic Cost-Insensitive Tasks. If the comprehensive service objective of a TN is $K = \{K_d > 0, K_e > 0, K_c = 0\}$, this task is not sensitive to the economic cost during the offloading service and tends to achieve low delay and high energy efficiency.

For this kind of tasks, the optimization goal of **P2** becomes $K_d D_{Oi} + K_e E_i$. This will not change the procedure of the SCOTT algorithm, which transforms the original problem into two one-variable form subproblems. Therefore, the SCOTT algorithm is applicable to economic cost-insensitive tasks. This case corresponds to the researches seeking the balance between task delay and energy cost, such as the DEBTS algorithm proposed in [51].

We omit the analyses of other cases like energy-sensitive ($K = \{K_d = 0, K_e > 0, K_c = 0\}$) tasks, which corresponds to the optimization problems in [41, 47]. The above investigations reveal the universal applicability of the SCOTT algorithm for multitarget tasks, and this will be further verified by the numerical simulation results in the next section.

7. Numerical Results

In this section, plenty of numerical simulations are carried out to investigate the performance of our proposed scheduling scheme. The task offloading solution and the corresponding service performance are evaluated for tasks with various service targets.

7.1. Simulation Setting. A heterogonous computing network is considered, in which HNs with various service capabilities are randomly distributed in the TN-centered computing network. The TN calls for offloading services from the task scheduler, which collect the service objectives of the TN and the service capabilities of the HNs. The offloading subtask is transmitted from the TN to the selected HN through a flat wireless channel with bandwidth of 10 MHz. The interference power I_i and the noise power spectral density N_0 are −43 dBm and −173 dBm/Hz, respectively. The path loss factor γ_i (in dB) is obtained through $38.46 + 20\log_{10}(d_i)$, where d_i (in m) is the distance between the TN and HN i. Besides, a shadowing factor −5 dB is adopted for each HN. The other parameter settings are specified in the corresponding simulation results.

Table 2: Capabilities of the TN and HN i.

Node	Parameter	Value
TN	η_T	1000 cycle/bit
	f_T	1 GHz
	θ_T	2×10^{-9} J/cycle
HN i	η_i	1000 cycle/bit
	f_i	10 GHz
	θ_i	1×10^{-9} J/cycle
	π_i	5×10^{-7} \$/bit

7.2. CQoS Maximal Offloading Solution. Firstly, we investigate the task offloading services through a specific HN. The service objective of a TN is $K = \{1, 1, 1\}$, and the task size is 2 Mbits. The capabilities of the TN and HN i are provided in Table 2.

Figure 2 shows the local maximal CQoS Q_i^* and the weighted service metrics including delay metric $K_d D_i$, energy consumption metric $K_e E_i$, and the economic cost metric $K_c C_i$. In Figure 1, we plot the local maximal CQoS of the offloading service provided by HN i, and the distance between the TN and HN i varies from 10 m to 100 m. With the increasing of the upper bound for the terminal transmission power, the local maximal CQoS increases from a lower bound to an upper bound. When $p_{\max} = 0$, the task has to be processed locally at the TN. Therefore, the lower bound in Figure 1 is the CQoS of local processing, i.e., $\underline{Q}$, which has been discussed in Section 6. The weighted metrics of local processing are shown in Figure 1 ($p_{\max} = 0$). A higher $p_{\max}$ helps achieve lower task delay and energy consumption metrics, but a higher economic cost metric at the same time, which is revealed by the numerical results in Figure 1. When $p_{\max}$ increases continually, the energy consumption of the offloading service can not continually decreases. Because the energy consumption for transforming 1 bit data, i.e., $p_i/(WB_i)$, is an increasing function of p_i. As a result, an upper bound of Q_i^* is achieved. The simulations results also show that an HN located close to the TN can achieve a higher CQoS, for the reason that the transmission energy consumption and transmission delay are lower when the distance is small. For an HN that is too far away, the cost performance of the offloading service is too low; thus, the local processing is adopted. For example, the HN that is 100 m away from the TN cannot provide service better than local processing.

In Figure 3, the CQoS maximal task offloading solutions through HNs located at different distances are plotted. As shown in Figure 3(a), the optimal transmission power p_i^* equals to $p_{\max}$ when $p_{\max}$ is small. With the increasing of $p_{\max}$, p_i^* reaches an upper bound, which corresponds to the upper bound of Q_i^* shown in Figure 2(a). We can also observe that the upper bound of p_i^* is small when the

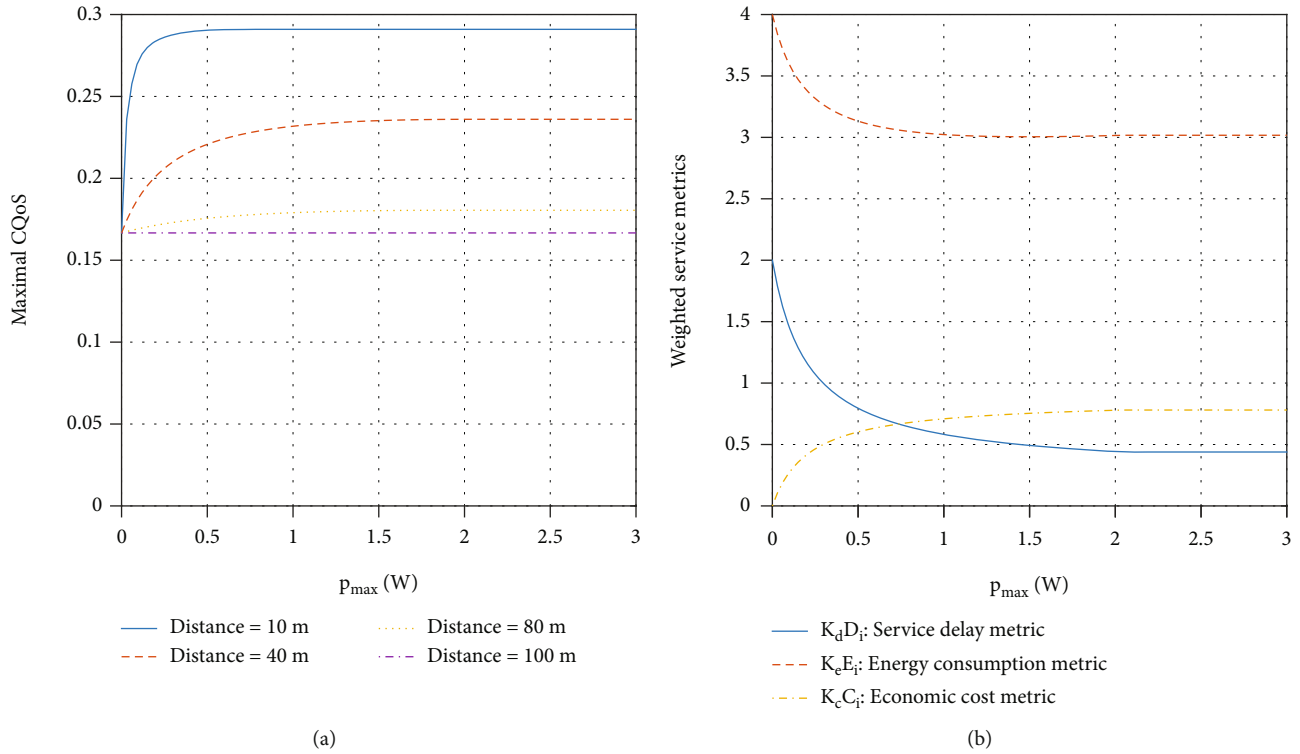

Figure 2: The maximal CQoS and the service metrics. (a) The maximal CQoS versus p_{max}. (b) The optimal service metrics versus p_{max}, $d_i = 40$ m.

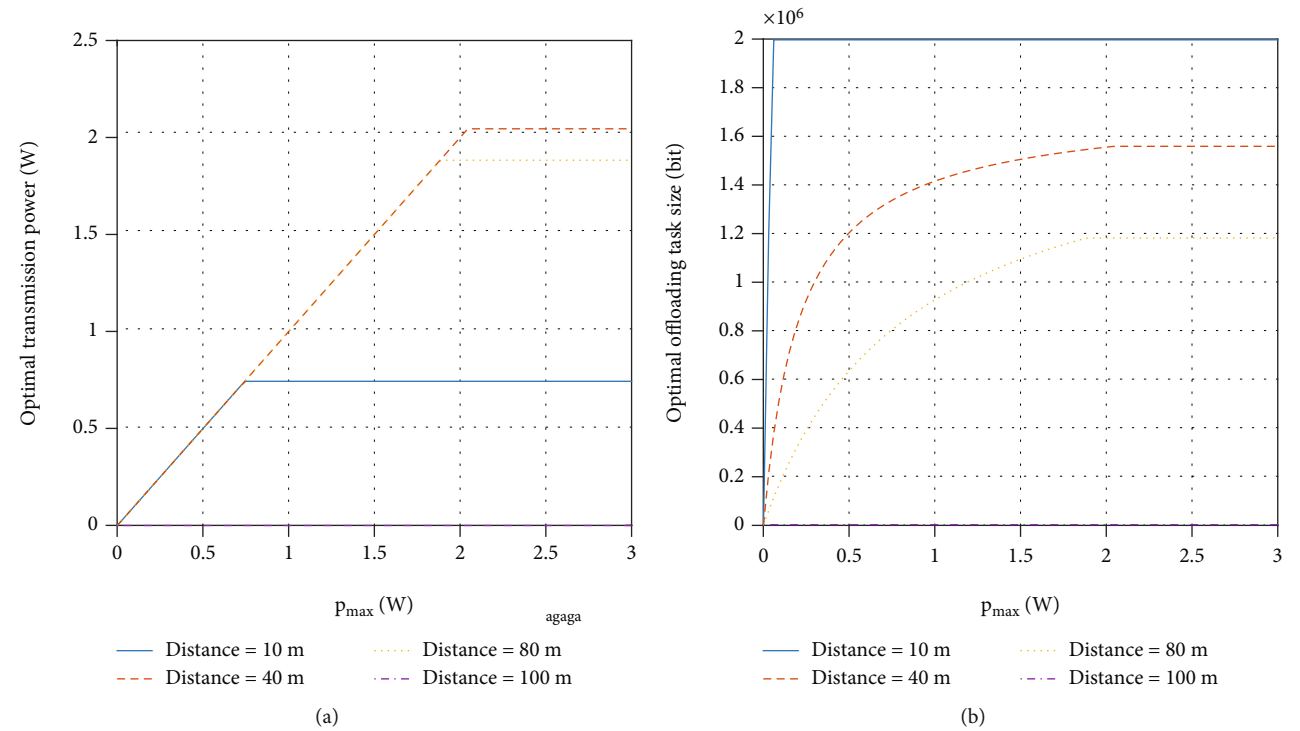

Figure 3: The CQoS maximal task offloading solutions through HNs located at different distances. (a) The optimal transmission power p_i^* versus p_{max}. (b) The optimal offloading task size l_i^* versus p_{max}.

Table 3: Capabilities of the available HNs.

Node	Capability				
	η_i	f_i	θ_i	π_i	d_i
HN 1	1000 cycle/bit	10 GHz	3×10^{-10} J/cycle	5×10^{-7} \$/bit	10 m
HN 2	1000 cycle/bit	5 GHz	1×10^{-10} J/cycle	5×10^{-7} \$/bit	20 m
HN 3	1000 cycle/bit	5 GHz	3×10^{-10} J/cycle	1.5×10^{-7} \$/bit	30 m

distance is small. It is because that the a small p_i can provide a high CQoS in this case. At the same time, the upper bound of p_i^* is also small when the distance is large. It is because that the increasing of p_i cannot provide a small task delay but a high energy consumption in this case. The extreme case in Figure 3(a) is the case when the distance is 100 m, which makes local processing be the optimal offloading solution. Figure 3(b) shows the optimal offloading task size l_i^* when $p_{\max}$ increases. We observe that the larger $p_{\max}$ is, the larger the subtask the TN offloads to the HN. For the nearby HNs with high cost performance, offloading the whole task can achieve the maximal CQoS. On the contrary, local processing is preferred when the HN is too far away, and the optimal offloading task size l_i^* is 0 in this case.

7.3. Offloading Service for Multitarget Tasks. Now, we investigate the task offloading service achieved by our proposed SCOTT algorithm, and three tasks with various service targets are considered. Specifically, the offloading service objectives of the three tasks are $\{1, 0, 0\}$, $\{0, 1, 0\}$, and $\{0.1,0.1,1\}$, respectively. The task size is 2 Mbits, and the upper bound of the transmission power is 4 W. Three HNs with various capabilities are available for the offloading service, and the parameters of the three HNs are provided in Table 3. Besides, the capabilities of the TN are the same with the previous subsection.

Figure 4 plots the maximal CQoS Q_i^* of the three tasks through the three available HNs. It is obvious that HN 1, HN 2, and HN 3 provide the highest CQoS for task 1, task 2, and task 3, respectively. Therefore, the SCOTT algorithm will assign task 1 to HN 1, and so on for the other two tasks.

It is easy to find that the three tasks are sensitive to different service metrics. Task 1 with service objective $\{1, 0, 0\}$ is a delay-sensitive task; thus, the SCOTT algorithm turns into the DOTS algorithm proposed in [26], which achieves the delay-minimized offloading scheme. Task 2 with service objective $\{0, 1, 0\}$ is an energy-sensitive task; thus, the SCOTT algorithm turns into the FEMTO algorithm proposed in [41], which achieves the energy-minimized offloading scheme. Task 3 with service objective $\{0.1,0.1,1\}$ is sensitive to service cost. The delay and energy factors of task 3 do not equal to 0 because that the service objective $\{0, 0, 1\}$ will lead to totally local processing and an infinitely great CQoS. On the other hand, Table 3 reveals that HN 1 has a fast CPU frequency, HN 2 has a low unit processing energy consumption, and HN 3 has a low unit economic cost. The above task and HN characteristics led to the numerical results in Figure 4. In conclusion, the SCOTT algorithm can provide the optimal offloading solution with the maximal CQoS for multitarget tasks, and it can universally cover the task scheduling schemes that concern limited metrics.

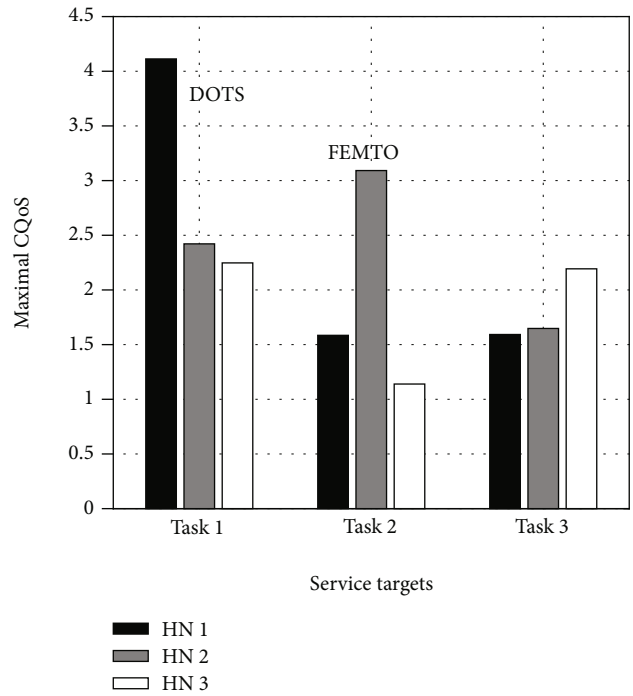

Figure 4: Maximal CQoS for 3-target tasks. The service objectives of the three tasks are $\{1, 0, 0\}$, $\{0, 1, 0\}$, and $\{0.1,0.1,1\}$, respectively. Task 1 and Task 2 can be scheduled with DOTS and FEMTO algorithm, respectively.

7.4. CQoS Maximal Offloading in Different Network Scenarios. Next, the CQoS maximal offloading through SCOTT algorithm in service clusters with different radius and HN amount is evaluated. Figure 5 plots the maximal CQoS of a TN with $K = \{1, 1, 1\}$. The HN amount equals to 10, 20, or 30, and the HNs are uniformly distributed in the service cluster with radius ranges from 20 m to 100 m. The parameters f_i, θ_i, and π_i of the HNs follows Gaussian distributions, the mean values of which are 5 GHz, 2×10^{-10} J/cycle, and 3×10^{-7} \$/bit, respectively.

We can observe from Figure 5 that the CQoS of the offloading service decreases with the increasing of the network radius. The reason is that the transmission delay and transmission energy consumption between the TN and the optimal HN increase with the increasing of the network radius, which lead to higher D_i and E_i. Besides, a large amount of available HNs can provide a higher CQoS for

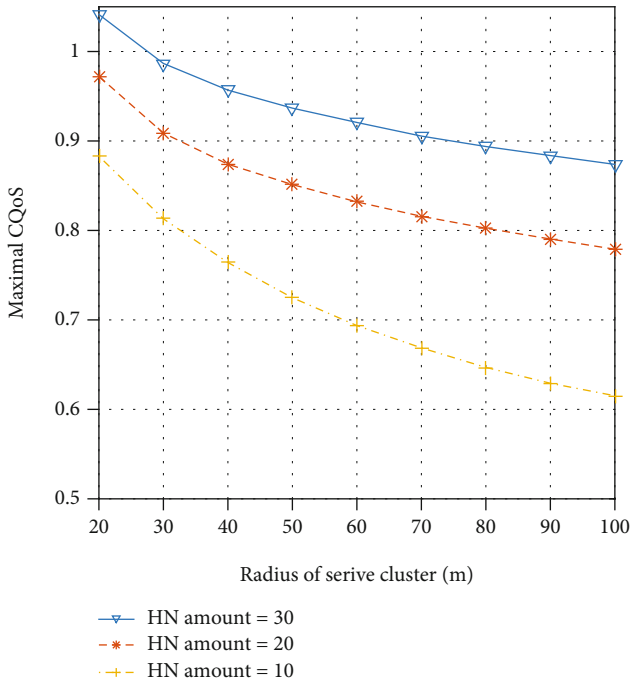

Figure 5: Maximal CQoS in service clusters with different radii and HN amount.

the TN. It is because that the probability to find a befitting HN for the TN with specific service objective increases when the HN amount is large.

8. Conclusions

In order to provide the customized services for the emerging multifarious IoT applications with multiple targets, we propose a general service framework in homogeneous MTMH computing networks, in which the TNs are served according to specific service objectives. The CQoS combining metrics including service delay, energy consumption, and economic cost is formulated to quantify the TN's comprehensive satisfactory level for the provided service. An algorithm named SCOTT is developed, which achieves the CQoS maximal offloading solutions by problem transforming. Numerical results based on extensive simulations in a heterogeneous computing network demonstrate that the proposed algorithm can effectively provide the optimal node selection, task division, and transmission power for the TNs with various service targets. The universal applicability of the task scheduling scheme is also verified by case studies and simulations. Future research directions of this work are the universal scheduling scheme for multitarget tasks of mobile network nodes in a multilayered computing network.

Appendix

A. Proof of Proposition 1

We prove Proposition 1 with contradiction. In other words, if the condition that $D_{Ti} \le D_{Oi}$ is not satisfied, the optimization goal Q_i in **P0** is not maximized.

Assume that there is an offloading scheme (l_i, p_i) for the available HN i, and the corresponding subtask delays satisfy $D_{Ti} > D_{Oi}$.

Based on this offloading scheme, we can decrease the TN transmission power p_i to $p_{i'} = p_i - \Delta_{p_i}$, such that $D_{Ti} = D_{Oi}$. Then, the delay target $K_d D_i = K_d \max(D_{Ti}, D_{Oi})$ and the economic cost target $K_c C_i = K_c l_i \pi_i$ remain unchanged.

The energy consumption for the transmission of 1 bit data, i.e.,

$$E_t = \frac{p_i}{W \log_2(1 + ((p_i \gamma_i \beta_i)/(I_i + WN_0)))}, 0 \le p_i \le p_{\max}, \tag{A.1}$$

is a monotone increasing function of the transmission power p_i when p_i is nonnegative [41]. So, the decrease of p_i leads to a small energy consumption target $K_e E_i = K_e(l_T \eta_T \theta_T + (l_i p_i / W B_i) + l_i \eta_i \theta_i)$.

Therefore, the utility $Q_i = 1/(K_d D_i + K_e E_i + K_c C_i)$ can always be increased by this transmission power adjustment, and Q_i is not maximized with the given offloading scheme (l_i, p_i).

The above proves Proposition 1.

B. Proof of Proposition 3

Take the derivative of $\partial Q_{i2}/\partial l_i$ with respect to p_i, we get

$$\frac{\partial^2 Q_{i2}}{\partial l_i \partial p_i} = \frac{1}{W B_i^2} \left[K_e B_i - \frac{(\gamma_i \beta_i)/(I_i + WN_0)}{(1 + ((p_i \gamma_i \beta_i)/(I_i + WN_0))) \ln 2} \cdot (K_d + K_e p_i) \right]. \tag{B.1}$$

Denote $K_e B_i - (\gamma_i \beta_i / (I_i + WN_0)/(1 + ((p_i \gamma_i \beta_i)/(I_i + WN_0))) \ln 2)(K_d + K_e p_i)$ by G. For nonnegative transmission power p_i, we have

$$\frac{\partial G}{\partial p_i} = K_e \frac{p_i (\gamma_i \beta_i / (I_i + WN_0))^2}{(1 + ((p_i \gamma_i \beta_i)/(I_i + WN_0))) \ln 2} + K_d \frac{(\gamma_i \beta_i / (I_i + WN_0))^2}{(1 + ((p_i \gamma_i \beta_i)/(I_i + WN_0)))^2 \ln 2} > 0,$$

$$\lim_{p_i \longrightarrow 0} G < 0,$$

$$\lim_{p_i \longrightarrow +\infty} G > 0. \tag{B.2}$$

Therefore, G is a monotone increasing function of p_i and has a single zero point when $p_i \in [0, \infty)$. Denote the only zero point of the equation $G = 0$ as $\widehat{p}_i$. Then, $\partial^2 Q_{i2}/\partial l_i \partial p_i$ is positive when $p_i \in [0, \widehat{p}_i]$ and negative when $p_i \in (\widehat{p}_i, \infty)$. In other words, the gradient $\partial Q_{i2}/\partial l_i$ is monotone decreasing when $p_i \in [0, \widehat{p}_i]$ and monotone increasing when $p_i \in (\widehat{p}_i, \infty)$.

The above proves Proposition 3.

C. Proof of Proposition 4

According to the expression of $\partial Q_{i2}/\partial l_i$ in (15), we can expand the objective function of **P2 − 1**, i.e., Q_{i21} in (18), as

$$\begin{aligned} Q_{i21} &= K_e l \eta_{\mathrm{T}} \theta_{\mathrm{T}} + \frac{l(\eta_{\mathrm{T}}/f_{\mathrm{T}})}{((\eta_{\mathrm{T}}/f_{\mathrm{T}}) + (1/WB_i) + (\eta_i/f_i))} \\ &\quad \cdot \left[K_d \left(\frac{1}{WB_i} + \frac{\eta_i}{f_i} \right) + K_e \left(\frac{p_i}{WB_i} + \eta_i \theta_i - \eta_{\mathrm{T}} \theta_{\mathrm{T}} \right) + K_c \pi_i \right] \\ &= K_e l \eta_{\mathrm{T}} \theta_{\mathrm{T}} + \frac{l(\eta_{\mathrm{T}}/f_{\mathrm{T}})(\partial Q_{i2}/\partial l_i)}{((\eta_{\mathrm{T}}/f_{\mathrm{T}}) + (1/WB_i) + (\eta_i/f_i))} \geq K_e l \eta_{\mathrm{T}} \theta_{\mathrm{T}}. \end{aligned} \tag{C.1}$$

The inequation in the last line of the above formula comes from the fact that $\partial Q_{i2}/\partial l_i \geq 0, p_i \in \mathscr{R}_1$.

According to Algorithm 1, there is '$p_i \in \mathscr{R}_1$ when $\mathscr{R}_1 \neq [0, p_{\max}]$, and '$p_i$ is one of the zero points of $\partial Q_{i2}/\partial l_i$. Therefore, the optimal transmission power that minimizes Q_{i21} is $p_{i1}^* =$ 'p_i when $\mathscr{R}_1 \neq [0, p_{\max}]$, and the minimum value of Q_{i21} is $K_e l \eta_{\mathrm{T}} \theta_{\mathrm{T}}$ in this case.

When $\mathscr{R}_1 = [0, p_{\max}]$, we can get the poles of Q_{i21} by searching the solutions of $\partial Q_{i21}/\partial p_i = 0$ in the value range $[0, p_{\max}]$ with bisection method. By comparing the values of Q_{i21} at the poles and $p_i = 0, p_i = p_{\max}$, we can get the optimal offloading solution in this case. Next, we can prove that the function $\partial Q_{i21}/\partial p_i = 0$ has at most one solution in $[0, p_{\max}]$, which is denoted by $\check{p}_i$.

The first derivative of Q_{i21} with respect to p_i is

$$\begin{aligned} \frac{\partial Q_{i21}}{\partial p_i} &= \frac{l(\eta_{\mathrm{T}}/f_{\mathrm{T}})}{((\eta_{\mathrm{T}}/f_{\mathrm{T}}) + (1/WB_i) + (\eta_i/f_i))^2} \\ &\quad \cdot \left\{ K_e \left(\frac{1}{WB_i} \right)^2 + K_e \left(\frac{\eta_{\mathrm{T}}}{f_{\mathrm{T}}} + \frac{\eta_i}{f_i} \right) \frac{1}{WB_i} \right. \\ &\quad + \left[K_d \frac{\eta_{\mathrm{T}}}{f_{\mathrm{T}}} - K_e(\eta_i \theta_i - \eta_{\mathrm{T}} \theta_{\mathrm{T}}) - K_c \pi_i \right. \\ &\quad \left. \left. + K_e \left(\frac{\eta_{\mathrm{T}}}{f_{\mathrm{T}}} + \frac{\eta_i}{f_i} \right) p_i \right] \frac{\partial (1/WB_i)}{\partial p_i} \right\}, \end{aligned} \tag{C.2}$$

in which $\partial(1/WB_i)/\partial p_i$ can be expanded as

$$\begin{aligned} \frac{\partial (1/WB_i)}{\partial p_i} &= \frac{-\gamma_i \beta_i / (I_i + WN_0)}{W[\log_2(1 + ((p_i \gamma_i \beta_i)/(I_i + WN_0)))]^2 (1 + ((p_i \gamma_i \beta_i)/(I_i + WN_0))) \ln 2} \\ &= \frac{-(\gamma_i \beta_i / (I_i + WN_0)) W}{(1 + (\gamma_i \beta_i / (I_i + WN_0)) p_i) \ln 2} \left(\frac{1}{WB_i} \right)^2. \end{aligned} \tag{C.3}$$

Therefore, the function $\partial Q_{i21}/\partial p_i = 0$ is equivalent to

$$\begin{aligned} &\left[K_d \frac{\eta_{\mathrm{T}}}{f_{\mathrm{T}}} - K_e(\eta_i \theta_i - \eta_{\mathrm{T}} \theta_{\mathrm{T}}) - K_c \pi_i + K_e \left(\frac{\eta_{\mathrm{T}}}{f_{\mathrm{T}}} + \frac{\eta_i}{f_i} \right) p_i \right] \\ &\cdot \frac{-(\gamma_i \beta_i / (I_i + WN_0)) W}{(1 + (\gamma_i \beta_i / (I_i + WN_0)) p_i) \ln 2} + K_e \\ &+ K_e \left(\frac{\eta_{\mathrm{T}}}{f_{\mathrm{T}}} + \frac{\eta_i}{f_i} \right) WB_i = 0. \end{aligned} \tag{C.4}$$

By introducing the variable substitution $x = 1 + (p_i \gamma_i \beta_i / (I_i + WN_0)) \in [1, +\infty)$, $\partial Q_{i21}/\partial p_i = 0$ can be further equivalent to

$$f(x) = \alpha_1 + \frac{\alpha_2}{x} + \alpha_3 \log_2(x) = 0, \tag{C.5}$$

in which the three parameters, i.e., α_1, α_2, and α_3, are

$$\begin{cases} \alpha_1 = -K_e \left(\dfrac{\eta_{\mathrm{T}}}{f_{\mathrm{T}}} + \dfrac{\eta_i}{f_i} \right) \dfrac{W}{\ln 2} + K_e, \\ \alpha_2 = \left\{ K_e \left(\dfrac{\eta_{\mathrm{T}}}{f_{\mathrm{T}}} + \dfrac{\eta_i}{f_i} \right) - \left[K_d \dfrac{\eta_{\mathrm{T}}}{f_{\mathrm{T}}} - K_e(\eta_i \theta_i - \eta_{\mathrm{T}} \theta_{\mathrm{T}}) - K_c \pi_i \right] \dfrac{\gamma_i \beta_i}{I_i + WN_0} \right\} \dfrac{W}{\ln 2}, \\ \alpha_3 = K_e \left(\dfrac{\eta_{\mathrm{T}}}{f_{\mathrm{T}}} + \dfrac{\eta_i}{f_i} \right) W > 0. \end{cases} \tag{C.6}$$

Now, we need to prove that the equation $f(x) = 0$ in (C.5) has at most one solution in $[1, 1 + (p_{\max}\gamma_i\beta_i/(I_i + WN_0))]$.

The first derivative of $f(x)$ with respect to x is

$$f'(x) = -\frac{\beta}{x^2} + \gamma\frac{1}{x\ln 2}. \tag{C.7}$$

Then, the solution of $f'(x) = 0$ in $(-\infty, +\infty)$ is $x_1 = \alpha_2 \ln 2/\alpha_3$.

If $x_1 < 1$, we have $f'(x) > 0$ when $x \geq 1$, and $f(x)$ is monotone increasing in $[1, 1 + (p_{\max}\gamma_i\beta_i/(I_i + WN_0))]$. Therefore, $f(x)$ has a single zero point in $[1, 1 + (p_{\max}\gamma_i\beta_i/(I_i + WN_0))]$ iff $f(1) < 0$ and $f(1 + (p_{\max}\gamma_i\beta_i/I_i + WN_0)) > 0$ are satisfied. The only zero point $\check{x} = 1 + (\check{p}_i\gamma_i\beta_i/(I_i + WN_0))$ can be obtained with bisection method.

If $x_1 \geq 1$, x_1 is the only pole of $f(x)$, and $f(x)$ is monotone decreasing in $[1, x_1]$ and monotone increasing in $(x_1, +\infty)$. Besides, we have

$$f(x_1) = \alpha_1 + \frac{\alpha_2}{\ln 2} + \alpha_3\log_2\left(\frac{\alpha_2\ln 2}{\alpha_3}\right), \tag{C.8}$$

in which $\alpha_1 + (\alpha_2/\ln 2) = K_e > 0$ and $\alpha_3\log_2(\alpha_2\ln 2/\alpha_3) \geq 0$. Thus, the minimum value of $f(x)$ is larger than 0, and $f(x) = 0$ has no solution in this case.

In conclusion, $p_{i1}^* = \text{`}p_i$ when $\mathscr{R}_1 \neq [0, p_{\max}]$. When $\mathscr{R}_1 = [0, p_{\max}]$, we get p_{i1}^* by comparing the values of Q_{i21} when $p_i \in \mathscr{P}_{i1}$, and the set $\mathscr{P}_{i1}$ includes the only pole of Q_{i21}, i.e, $\check{p}_i$ iff the three conditions in (22) are satisfied.

The above proves Proposition 4.

D. Proof of Proposition 5

The first derivative of Q_{i22} with respect to p_i is

$$\frac{\partial Q_{i22}}{\partial p_i} = \frac{l}{WB_i^2}\left[K_eB_i - \frac{\gamma_i\beta_i/(I_i + WN_0)}{(1 + (p_i\gamma_i\beta_i/(I_i + WN_0)))\ln 2} \cdot (K_d + K_ep_i)\right]. \tag{D.1}$$

Compare (D.1) with (B.1), we get

$$\frac{\partial Q_{i22}}{\partial p_i} = l\frac{\partial^2 Q_{i2}}{\partial l_i\partial p_i}. \tag{D.2}$$

Therefore, the variation trend of Q_{i22} is the same with $\partial Q_{i2}/\partial l_i$. Based on the conclusions in Proposition 3 and Algorithm 1, we can conclude that Q_{i22} is monotone decreasing when $p_i \in [0, \widehat{p}_i]$ and monotone increasing when $p_i \in (\widehat{p}_i, \infty)$.

For the cases that $\partial Q_{i2}/\partial l_i|_{\widehat{p}_i} \geq 0$ or $p_{\max} \leq \text{`}p_i$, $\mathscr{R}_2 = \varnothing$. Otherwise, we have the following conditions:

If $\text{`}p_i < p_{\max} < \widehat{p}_i$, $\mathscr{R}_2 = [\text{`}p_i, p_{\max}] \neq \varnothing$, and Q_{i22} is monotone decreasing in $\mathscr{R}_2$, thus $p_{i2}^* = p_{\max}$.

If $\widehat{p}_i \leq p_{\max} < p'_i$, $\mathscr{R}_2 = [\text{`}p_i, p_{\max}] \neq \varnothing$, and $Q_{i22}$ is monotone decreasing in $[\text{`}p_i, \widehat{p}_i]$ and monotone increasing in $[\widehat{p}_i, p_{\max}]$, thus $p_{i2}^* = \widehat{p}_i$.

If $p_{\max} \geq p'_i$, $\mathscr{R}_2 = [\text{`}p_i, p'_i] \neq \varnothing$, and $Q_{i22}$ is monotone decreasing in $[\text{`}p_i, \widehat{p}_i]$ and monotone increasing in $[\widehat{p}_i, p'_i]$, thus $p_{i2}^* = \widehat{p}_i$.

The corresponding optimal subtask size l_{i2}^* equals to l for the reason that Q_{i2} is a monotone decreasing function of l_i when $p_i \in \mathscr{R}_2$.

We have proved the optimal solution of **P2 − 2** so far. Then, we need to prove that $Q_{i21}^* \geq Q_{i22}^*$ when $\mathscr{R}_2 \neq \varnothing$.

In the proof of Proposition 4, we express Q_{i21} as

$$Q_{i21} = K_el\eta_T\theta_T + \frac{l(\eta_T/f_T)(\partial Q_{i2}/\partial l_i)}{((\eta_T/f_T) + (1/WB_i) + (\eta_i/f_i))} \geq K_el\eta_T\theta_T. \tag{D.3}$$

Besides, Q_{i21} can be expressed as

$$\begin{aligned} Q_{i22} &= K_dl\left(\frac{1}{WB_i} + \frac{\eta_i}{f_i}\right) + K_el\left(\frac{p_i}{WB_i} + \eta_i\theta_i\right) + K_cl\pi_i \\ &= K_el\eta_T\theta_T + l\frac{\partial Q_{i2}}{\partial l_i} < K_el\eta_T\theta_T, \end{aligned} \tag{D.4}$$

in which the in inequality in the last line of the above formula comes from the fact that $\partial Q_{i2}/\partial l_i \geq 0, p_i \in \mathscr{R}_2$.

Therefore, Q_{i21} is always larger than Q_{i22}, and there is $Q_{i21}^* \geq Q_{i22}^*$ when $\mathscr{R}_2 \neq \varnothing$.

The above proves Proposition 5.

Data Availability

No data were used to support this study.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was partially supported by the Natural Science Foundation of Shandong Province, China, under grant ZR2021QF090, partially by the National Key Research and Development Program of China under grant 2020YFB2104300, and partially by the National Natural Science Foundation of China (NSFC) Key Project Program under grant 61932014.

References

[1] M. Agiwal, A. Roy, and N. Saxena, "Next generation 5G wireless networks: a comprehensive survey," *IEEE Communications Survey & Tutorials*, vol. 18, no. 3, pp. 1617–1655, 2016.

[2] G. A. Akpakwu, B. J. Silva, G. P. Hancke, and A. M. Abu-Mahfouz, "A survey on 5G networks for the internet of things: communication technologies and challenges," *IEEE Access*, vol. 6, pp. 3619–3647, 2018.

[3] S. Li, L. D. Xu, and S. Zhao, "The Internet of Things: a survey," *Information Systems Frontiers*, vol. 17, no. 2, pp. 243–259, 2015.

[4] A. Botta, W. D. Donato, and V. Persico, "Integration of Cloud computing and Internet of Things: a survey," *Future Generation Computer Systems*, vol. 56, no. C, pp. 684–700, 2016.

[5] S. Singh, A. Singh, and V. Goyal, "Cloud of things: a systematic review on issues and challenges in integration of cloud computing and internet of things," in *Recent Innovations in Computing*, pp. 573–587, Springer, Singapore, 2021.

[6] X. Chen, L. Jiao, W. Li, and X. Fu, "Efficient multi-user computation offloading for mobile-edge cloud computing," *IEEE/ACM Transactions on Networking*, vol. 24, no. 5, pp. 2795–2808, 2016.

[7] M. Chiang, S. Ha, I. Chih-Lin, F. Risso, and T. Zhang, "Clarifying fog computing and networking: 10 questions and answers," *IEEE Communications Magazine*, vol. 55, no. 4, pp. 18–20, 2017.

[8] Y. Yang, J. Xu, G. Shi, and C.-X. Wang, *5G Wireless Systems: Simulation and Evaluation Techniques*, Springer, 2017.

[9] Y. Yu, "Mobile edge computing towards 5G: vision, recent progress, and open challenges," *China Communications*, vol. 13, no. 2, pp. 89–99, 2016.

[10] Y. J. Ku, D. Y. Lin, C. F. Lee et al., "5G radio access network design with the fog paradigm: confluence of communications and computing," *IEEE Communications Magazine*, vol. 55, no. 4, pp. 46–52, 2017.

[11] Y. Yang, "Multi-tier computing networks for intelligent IoT," *Nature Electronics*, vol. 2, no. 1, pp. 4-5, 2019.

[12] J. Ni, K. Zhang, X. Lin, and X. S. Shen, "Securing fog computing for Internet of Things applications: challenges and solutions," *IEEE Communications Surveys Tutorials*, vol. 20, no. 1, pp. 601–628, 2018, 20.

[13] Y. Yang, Z. Liu, X. Yang, K. Wang, X. Hong, and X. Ge, "POMT: paired offloading of multiple tasks in heterogeneous fog networks," *IEEE Internet of Things Journal*, vol. 6, no. 5, pp. 8658–8669, 2019.

[14] W. Shi, J. Cao, Q. Zhang, Y. Li, and L. Xu, "Edge computing: vision and challenges," *IEEE Internet of Things Journal*, vol. 3, no. 5, pp. 637–646, 2016.

[15] N. A. Cv and R. Lavanya, "Fog computing and its role in the Internet of Things," in *Advancing Consumer-Centric Fog Computing Architectures*, pp. 63–71, IGI Global, 2019.

[16] S. Sthapit, J. Thompson, N. M. Robertson, and J. R. Hopgood, "Computational load balancing on the edge in absence of cloud and fog," *IEEE Transactions on Mobile Computing*, vol. 18, no. 7, pp. 1499–1512, 2019.

[17] J. Dizdarevic, F. Carpio, A. Jukan, and X. Masip-Bruin, "Survey of communication protocols for internet-of-things and related challenges of fog and cloud computing integration," *ACM Computing Surveys*, vol. 51, no. 6, pp. 1–29, 2018.

[18] T. Qiu, J. Chi, X. Zhou, Z. Ning, M. Atiquzzaman, and D. O. Wu, "Edge computing in industrial internet of things: architecture, advances and challenges," *IEEE Communications Surveys Tutorials*, vol. 22, no. 4, pp. 2462–2488, 2020.

[19] M. Baidas, "Resource allocation for offloading-efficiency maximization in clustered NOMA- enabled mobile edge computing networks," *Computer Networks*, vol. 189, p. 107919, 2021.

[20] J. Liu, Y. Mao, J. Zhang, and K. B. Letaief, "Delay-optimal computation task scheduling for mobile-edge computing systems," in *IEEE International Symposium on Information Theory (ISIT)*, pp. 1451–1455, Barcelona, Spain, July 2016.

[21] X. Zhang and T. Wang, "Elastic and reliable bandwidth reservation based on distributed traffic monitoring and control," *IEEE Transactions on Parallel and Distributed Systems*, vol. 33, no. 12, pp. 4563–4580, 2022.

[22] X. Zhang, Y. Wang, G. Geng, and J. Yu, "Delay-optimized multicast tree packing in software-defined networks," *IEEE Transactions on Services Computing*, vol. 1, pp. 1–14, 2021.

[23] Z. Ning, J. Huang, and X. Wang, "Vehicular fog computing: enabling real-time traffic management for smart cities," *IEEE-Wireless Communications*, vol. 26, no. 1, pp. 87–93, 2019.

[24] A. A. Khadir and S. A. H. Seno, "SDN-based offloading policy to reduce the delay in fog-vehicular networks," *Peer-to-Peer Networking and Applications*, vol. 14, no. 3, pp. 1261–1275, 2021.

[25] M.-H. Chen, M. Dong, and B. Liang, "Resource sharing of a computing access point for multi-user mobile cloud offloading with delay constraints," *IEEE Transactions on Mobile Computing*, vol. 17, no. 12, pp. 2868–2881, 2018.

[26] G. Zhang, F. Shen, N. Chen, P. Zhu, X. Dai, and Y. Yang, "DOTS: delay-optimal task scheduling among voluntary nodes in fog networks," *IEEE Internet of Things Journal*, vol. 6, no. 2, pp. 3533–3544, 2019.

[27] A. Yousefpour, G. Ishigaki, and J. P. Jue, "Fog computing: towards minimizing delay in the Internet of Things," in *2017 IEEE International Conference on Edge Computing (EDGE)*, pp. 17–24, Honolulu, HI, USA, Jun. 2017.

[28] Z. Liu, Y. Yang, K. Wang, Z. Shao, and J. Zhang, "POST: parallel offloading of splittable tasks in heterogeneous fog networks," *IEEE Internet of Things Journal*, vol. 7, no. 4, pp. 3170–3183, 2020.

[29] S. Zhang, N. Zhang, S. Zhou, J. Gong, Z. Niu, and X. Shen, "Energy-aware traffic offloading for green heterogeneous networks," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 5, pp. 1116–1129, 2016.

[30] F. Jalali, K. Hinton, R. Ayre, T. Alpcan, and R. S. Tucker, "Fog computing may help to save energy in cloud computing," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 5, pp. 1728–1739, 2016.

[31] J. Kwak, Y. Kim, J. Lee, and S. Chong, "DREAM: dynamic resource and task allocation for energy minimization in mobile cloud systems," *IEEE Journal on Selected Areas in Communications*, vol. 33, no. 12, pp. 2510–2523, 2015.

[32] D. Zhang, Z. Chen, M. K. Awad, N. Zhang, H. Zhou, and X. S. Shen, "Utility-optimal resource management and allocation algorithm for energy harvesting cognitive radio sensor networks," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 12, pp. 3552–3565, 2016.

[33] B. Li, Z. Fei, J. Shen, X. Jiang, and X. Zhong, "Dynamic offloading for energy harvesting mobile edge computing: architecture, case studies, and future directions," *IEEE Access*, vol. 7, pp. 79877–79886, 2019.

[34] S. Guo, J. Liu, Y. Yang, B. Xiao, and Z. Li, "Energy-efficient dynamic computation offloading and cooperative task scheduling in mobile cloud computing," *IEEE Transactions on Mobile Computing*, vol. 18, no. 2, pp. 319–333, 2019.

[35] K. Kaur, T. Dhand, N. Kumar, and S. Zeadally, "Container-as-a-service at the edge: trade-off between energy efficiency and service availability at fog nano data centers," *IEEE Wireless Communications*, vol. 24, no. 3, pp. 48–56, 2017.

[36] S. Zhou, W. Jadoon, and J. Shuja, "Machine learning-based offloading strategy for lightweight user mobile edge computing tasks," *Complexity*, vol. 2021, Article ID 6455617, 11 pages, 2021.

[37] X. Chen, C. Tang, Z. Li, L. Qi, S. Chen, and S. Chen, "A pricing approach toward incentive mechanisms for participant mobile crowdsensing in edge computing," *Mobile Networks and Applications*, vol. 25, no. 4, pp. 1220–1232, 2020.

[38] W. Hu and G. Cao, "Quality-aware traffic offloading in wireless networks," *IEEE Transactions on Mobile Computing*, vol. 16, no. 11, pp. 3182–3195, 2017.

[39] F. Shen, G. Zhang, C. Zhang, Y. Yang, and R. Yang, "An incentive framework for resource sensing in fog computing networks," in *IEEE Global Communications Conference (GLOBECOM)*, pp. 1–6, Abu Dhabi, United Arab Emirates, December 2018.

[40] D. Kim, H. Lee, H. Song, N. Choi, and Y. Yi, "Economics of fog computing: interplay among infrastructure and service providers, users, and edge resource owners," *IEEE Transactions on Mobile Computing*, vol. 19, no. 11, pp. 2609–2622, 2020.

[41] G. Zhang, F. Shen, Z. Liu, Y. Yang, K. Wang, and M. Zhou, "FEMTO: fair and energy-minimized task offloading for fog-enabled IoT networks," *IEEE Internet of Things Journal*, vol. 6, no. 3, pp. 4388–4400, 2019.

[42] S. N. Shirazi, A. Gouglidis, A. Farshad, and D. Hutchison, "The extended cloud: review and analysis of mobile edge computing and fog from a security and resilience perspective," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 11, pp. 2586–2595, 2017.

[43] A. Aral and I. Brandic, "Learning spatiotemporal failure dependencies for resilient edge computing services," *IEEE Transactions on Parallel and Distributed Systems*, vol. 32, no. 7, pp. 1578–1590, 2021.

[44] Y. Sun and N. Zhang, "A resource-sharing model based on a repeated game in fog computing," *Saudi Journal of Biological Sciences*, vol. 24, no. 3, pp. 687–694, 2017.

[45] H. Zhang, Y. Zhang, Y. Gu, D. Niyato, and Z. Han, "A hierarchical game framework for resource management in fog computing," *IEEE CommunicationsMagazine*, vol. 55, no. 8, pp. 52–57, 2017.

[46] R. A. C. da Silva and N. L. S. da Fonseca, "Location of fog nodes for reduction of energy consumption of end-user devices," *IEEE Transactions on Green Communications and Networking*, vol. 4, no. 2, pp. 593–605, 2020.

[47] Y. Yang, K. Wang, G. Zhang, X. Chen, X. Luo, and M.-T. Zhou, "MEETS: maximal energy efficient task scheduling in homogeneous fog networks," *IEEE Internet of Things Journal*, vol. 5, no. 5, pp. 4076–4087, 2018.

[48] T. Zhao, S. Zhou, L. Song, Z. Jiang, X. Guo, and Z. Niu, "Energy-optimal and delay-bounded computation offloading in mobile edge computing with heterogeneous clouds," *China Communications*, vol. 17, no. 5, pp. 191–210, 2020.

[49] E. Mustafa, J. Shuja, A. I. Jehangiri et al., "Joint wireless power transfer and task offloading in mobile edge computing: a survey," *Cluster Computing*, vol. 25, no. 4, pp. 2429–2448, 2022.

[50] A. I. Jehangiri, T. Maqsood, A. I. Umar et al., "LiMPO: lightweight mobility prediction and offloading framework using machine learning for mobile edge computing," *Cluster Computing*, pp. 1–19, 2022.

[51] Y. Yang, S. Zhao, W. Zhang, Y. Chen, X. Luo, and J. Wang, "DEBTS: delay energy balanced task scheduling in homogeneous fog networks," *IEEE Internet of Things Journal*, vol. 5, no. 3, pp. 2094–2106, 2018.