Research on Joint Optimization of Edge Computer Offload Strategy and Wireless Resource Allocation

HuiNan Wang1,*, JiaYi Liu1, JianMin Zhang1

1State Grid Shanxi Marketing Service Center, Taiyuan, Shanxi, China, 030000

*Corresponding author e-mail: 15333664663@189.com

Abstract. To solve the problem that users cannot use network resources efficiently, based on mobile edge computer network architecture, in a resource-constrained system, an energy-efficient offload strategy and a wireless resource joint optimization algorithm (EEJS) are proposed to achieve a balance between user diversity and MS diversity.

Keywords: MEC, Uninstall Strategy, Wireless Resources, Allocation Optimization

1. Introduction

The explosive growth of terminal devices, especially smartphones and Internet of Things devices, has led to increasing user demand for low-latency, high-quality services such as mobile games and augmented reality/virtual reality (AR/VR). At the same time, the demand for network resources, especially computing resources, of these emerging services is increasing, so that user equipment cannot be executed and processed locally. Cloud computing is well known for its powerful computing capabilities. However, because cloud resources are often very far away from edge users in terms of spatial distance, the data transmission delay is very large, which cannot meet the strict requirements of computing services for low latency. A large amount of business data at the edge of the network flocks to the cloud, which is prone to network link congestion and increase the load on the core network. In the Mobility Edgability Computing (MEC) system, the diversity of users and MEC Server (MS), the coupling relationship between the transmission and calculation processes in terms of delay constraints, and the volatility and difference of channels all make the optimization of network resource allocation and task offloading strategies difficult. In order to optimize the overall energy consumption of the system, this paper proposes a joint optimization algorithm for offload strategy and wireless resource allocation in a multi-user multi-MS network. Using the two-layer optimization algorithm as a guide, the original NP-hard problem is flexibly divided into two sub-optimization problems, namely, lower-level and upper-level, which effectively reduces the complexity of the algorithm while achieving the optimization effect. In a complex network scenario with multiple users and multiple
MEC servers, the overall energy consumption of the system is effectively reduced. Especially in a relatively sparse network scenario, the optimization effect is more obvious.

2. Basic overview of MEC

As a network technology paradigm built on a virtualization platform, MEC is regarded as a marginal evolution of cloud computing. It collects idle computing and storage resources distributed at the edge of the network to perform user-intensive offload. The type of computing task required. The MEC system uses NetFunctions virtualization(NFV), Information-Centrical NetWorks(ICN) and software-defined networks (SDN) to create a large number of simultaneous network devices, which enables virtual network devices such as multiple mobile devices. For computing services, ICN provides an alternative end-to-end service identification paradigm for MEC networks, enabling information-centric content-aware computing. SDN technology enables the MEC network to perform functional abstractions, thereby achieving dynamic and scalable computing.

3. System model

As shown in Figure 1, multiple users and multiple MSs form a MEC network system. Orthogonal Frequently Diversified Multiple Access (OFDMA) acts as an upstream data transmission mechanism. Multiple MSs with certain computing capabilities are deployed on the base station (BS) side of the base station to provide the ability to perform tasks. It is assumed that the MS works independently and is connected to the core network through optical fiber. At the same time, it is assumed that the MS is configured with Central Processing Unit (CPU) and is currently idle, and $K$ MSs are represented as set $\kappa=\{1,2,\cdots, K\}$. N subcarriers with the same bandwidth and size $B_N$ are used for uplink wireless transmission, and are represented as set $\eta=\{1,2,\cdots, N\}$. There are $I$ users at the edge of the network, which is represented as set $T=\{1,2,\cdots, I\}$. Each user has a task to perform, that is, there are $I$ independent tasks in the network. In order not to cause ambiguity, $i$ is also used as the indicator of the task. The attributes of each task are described by three parameters $A(D_i, \tau_i, X_i)$, where $D_i$ indicates the task's input data size (in bits), $\tau_i$ indicates the task completion delay constraint (in seconds), and $X_i$ indicates the task's computational intensity (In units of CPU cycle per bit). The attribute parameters of these tasks are only related to the nature of the task itself. In addition, assuming that MS has sufficient calculation and execution capabilities, once the offloaded task is received, it will continue to execute until the task is completed.
The local CPU frequency of user $i$ is expressed as $f_{i,\text{loc}}$, the maximum transmission power of each user is the same and is set to $P^m$, the CPU frequency of MS $k$ is expressed as $f_{k,\text{ser}}$, and the allocation matrix of subcarriers is expressed as
\[ W = \left\{ w_{i,n,k} \bigg| w_{i,n,k} \in \{0,1\}, i \in T, n \in \eta, k \in \kappa \right\}, \]
where $w_{i,n,k} = 1$ indicates subcarrier $n$ assigned to user $i$ and user $i$'s tasks are offloaded to MS $k$ for execution, $w_{i,n,k} = 0$ means the opposite. The power allocation matrix of the subcarriers is expressed as
\[ P = \left\{ p_{i,n,k} \bigg| p_{i,n,k} \in \{0,P^m\}, i \in T, n \in \eta, k \in \kappa \right\}, \]
where $p_{i,n,k}$ represents the transmission power allocated to the subcarrier $n$ of the user $i$ when the user $i$ offloads the task to the MS for execution. The channel gain matrix on the subcarrier is represented by
\[ G = \left\{ g_{i,n,k} \bigg| i \in T, n \in \eta, k \in \kappa \right\}, \]
where $g_{i,n,k}$ represents the channel gain between the user $i$ and MS $k$ on the subcarrier $n$. In addition, it is assumed that the system is in a stable channel fading environment, so that the channel gain matrix $G$ can be kept constant during a task scheduling process. The noise of the system follows a Gaussian distribution with an expected value of 0, which is expressed as $\delta^2$.

4. Optimization problem reconstruction and analysis

For the multi-user multi-MS network scenario, consider whether to pre-determine whether the user offloads the task, so that the user's tasks in the network can be divided into two categories: local execution and offload execution, and then further offload strategies for the tasks that need to be offloaded. Joint optimization of wireless resources. In this section, we will propose a strategy for task execution mode determination, and at the same time build and analyze system optimization problems.

4.1. Task execution mode decision strategy
The division of task execution modes is based on energy consumption and delay. In order to avoid excessive energy consumption on user terminal devices, the upper threshold of local execution energy consumption is set to $E_0$, taking into account the user’s local execution. The time cost cannot exceed the delay constraint of the task itself, the specific division strategy can be expressed as 1-1.

$$\begin{align*}
    b_{i,0} = 1: \{E_i < E_0\} \cap \{T_i < \tau_i\} \\
    b_{i,0} = 0: \text{other situations}
\end{align*}$$

(1-1)

Among them, $b_{i,0} = 1$ means that task $i$ is executed locally. At this time, the condition that needs to be met is that the energy consumption of the user’s local execution is less than the threshold $E_0$ and the execution time cost does not exceed the task execution delay constraint $\tau_i$; $b_{i,0} = 0$ means the task $i$ should be uninstalled and executed on MS. When the task is offloaded to the MS for execution, an optimization algorithm for offloading strategy and wireless resource joint allocation is further proposed.

4.2. The problem of minimizing total system energy consumption

According to the task execution mode decision strategy, the tasks of users in the entire network are divided into two types: cost execution and offload execution, so that tasks that need to be offloaded and executed in the network are updated to sets $T' = \{1, 2, \ldots, I'\}$ and $T' \subset T$. Based on the optimization problem of minimizing system energy consumption, a joint optimization problem of offload strategy and wireless resources is constructed, which is denoted as $\rho$.

$$\rho: \begin{align*}
    & \min F(b, W, P) = \sum_{i \in T'} \sum_{k = 1}^{K} b_{i,k} E_{i,k}^r \\
    & \text{s.t.} C1: b_{i,k} \in \{0, 1\}, \forall i \in T', \forall k \in \kappa \\
    & C2: \sum_{i \in T'} \sum_{k = 1}^{K} b_{i,k} \leq 1, \forall k \in \kappa \\
    & C3: \sum_{k = 1}^{K} b_{i,k} = 1, \forall i \in T' \\
    & C4: \sum_{k = 1}^{K} \sum_{n = 1}^{N} w_{i,n,k} p_{i,n,k} \leq P_i, \forall i \in T' \\
    & C5: w_{i,n,k} \in \{0, 1\}, \forall i \in T', \forall n \in \eta, \forall k \in \kappa \\
    & C6: \sum_{i \in T'} \sum_{k = 1}^{K} w_{i,n,k} = 1, \forall n \in \eta \\
    & C7: T = \sum_{k = 1}^{K} b_{i,k} \left( \frac{D}{R_{i,k}} + \frac{D_X}{f_{k,ser}} \right) \leq \tau_i, \forall i \in T'
\end{align*}$$

(1-2)
Among them, 
\[ b = \left\{ b_{i,k}, \forall i \in T', \forall k \in \kappa \right\}, \quad W = \left\{ w_{i,n,k}, \forall i \in T', \forall n \in \eta, \forall k \in \kappa \right\}, \]

Constraint condition \( C1 \) indicates that the uninstall strategy \( b_{i,k} \) is a binary variable used to indicate whether the task \( i \) is uninstalled to the MS \( k \). \( C2 \) means that MS only receives no more than one task at a time. Constraint \( C3 \) means that each task is only offloaded to one MS. \( C4 \) represents the upper limit of the transmission power of each user. In addition, \( C5, C6 \) indicates that the subcarrier allocation strategy is a binary variable, and each subcarrier is only allocated to one user for wireless transmission with MS. \( C7 \) indicates that the execution of each task has strict delay constraints, that is, the cumulative time overhead of the transmission and calculation process cannot exceed the delay constraints of the task itself.

Obviously, resource allocation problem \( \rho \) is a mixed integer nonlinear programming problem. This problem is usually an NP-hard problem and cannot be solved directly. In addition, dealing with a joint optimization problem in which both the objective and constraint conditions are non-convex functions makes the solution of the entire problem extra difficult. Based on the multi-objective hierarchical optimization idea, the original problem \( \rho \) can be equivalently converted to \( \rho^1 \), as follows:

\[
\rho^1: \min_{b,W,P} F(b,W,P) = \sum_{i \in T} \sum_{k=1}^{\kappa} b_{i,k} E_{i,k}^r \\
\text{s.t. } C1 - C7 \\
C8: (W,P) \in \arg \min_{b \in B} F(b,W,P)
\]  \hspace{1cm} (1-3)

Where \( B \) represents the feasible domain of offload strategy \( b \), \( \rho^1 \) contains two nesting problems, on the one hand, only when the offload strategy is determined, the energy consumption of the entire system can be calculated, on the other hand, the offload strategy will be affected by the subcarrier and its transmission power allocation strategy. The impact of this model on the mutual constraints of multiple objectives is consistent with the modeling characteristics of the bi-level optimization problem. In view of the above reasons, the idea of hierarchical optimization is used, and the idea of bi-level optimization algorithm is used to deal with the optimization problem \( \rho^1 \).

5. Offloading strategy and wireless resource joint optimization strategy

Based on the problem \( \rho^1 \), consider using the bi-level optimization method to deal with the original optimization problem \( \rho \). Based on the hierarchical structure of \( \rho^1 \), the original problem is divided into two sub-optimization problems, lower-level and upper-level. According to the given offload strategy \( b \) in the lower-level problem, the optimal subcarrier allocation strategy \( W^* \) and transmission power allocation strategy \( P^* \) can be obtained by solving the objective function \( F(b,W,P) \); in the upper-level problem, according to the lower-level problem The optimal subcarrier
allocation strategy and its transmission power allocation strategy $P'$ obtained in the above, and the optimal offload strategy $b^*$ can be obtained by optimizing the objective function $F(b,W^*,P')$. This section will expand in detail from two aspects: the optimization of subcarriers and their power allocation strategies for the lower-level problem, and the search for the optimal offload strategy for the upper-level problem.

5.1. Subcarrier and optimization of its power allocation strategy

Given the task offload strategy $b$, the overall computational efficiency becomes a known quantity, and such a result can be simplified into problem $\rho^2$.

$$\rho^2 : \min_{W,P} F(b,W,P) = \sum_{i \in T} \sum_{k=1}^{K} b_{i,k} \sum_{n=1}^{n} w_{i,n,k} \frac{D_i}{R_{i,k}}$$

s.t. $C4 - C6$

$$C7 : T_i = \sum_{k=1}^{K} b_{i,k} \left( \frac{D_i}{R_{i,k}} + \frac{D_i X_i}{f_{k,ser}} \right) \leq \tau_i, \forall i \in T'$$

(1-4)

The constraint $C7$ can be further transformed into:

$$C9 : \sum_{k=1}^{K} b_{i,k} \frac{D_i}{R_{i,k}} \leq \chi_i, \chi_i = \tau_i - \sum_{k=1}^{K} b_{i,k} \frac{D_i X_i}{f_{k,ser}}, \forall i \in T'$$

(1-5)

Since the offload strategy $b_{i,k}$ is a binary variable, at the same time, according to the constraint condition $C3$, $C9$ can be further converted into:

$$C10 : b_{i,k} \frac{D_i}{R_{i,k}} \leq \chi_i, \forall i \in T', \forall k \in \kappa$$

(1-6)

Since the optimization problem and the constraint conditions are both non-convex functions and cannot be solved directly, the $\frac{(b_{i,k} D_i)}{R_{i,k}}$ term in the objective function $F'(b,W,P)$ is replaced with its upper limit $b_{i,k} \chi_i$ by using the constraint condition $C10$, so we get the convex approximation problem $\rho^2 - 1$ of problem $\rho^2$, as follows:

$$\rho^2 - 1 : \min_{W,P} \zeta(b,W,P) = \sum_{i=1}^{K} \sum_{k=1}^{K} b_{i,k} \sum_{n=1}^{n} w_{i,n,k} p_{i,n,k}$$

s.t. $C4 - C6, C10$

(1-7)
Obviously $\rho^{2-1}$ is a nearly strict convex optimization problem except for the discrete binary variable $w_{i,n,k}$, so the discrete variable $w_{i,n,k}$ is relaxed and transformed into a continuous variable belonging to the interval $[0,1]$. $C10$ can be further converted to:

$$C11: b_{i,k} \frac{D_i}{\chi_i} \leq R_{i,k}, \forall i \in T', \forall k \in \kappa$$

(1-8)

For the optimization of $\rho^{2-1}$, the Lagrangian multiplier method is used. First, the Lagrangian function is obtained according to the objective function and the constraints, as follows:

$$\mathcal{R}(\alpha, \beta, W, P) = \sum_{i \in T} \sum_{k=1}^{K} b_{i,k} \sum_{n=1}^{N} w_{i,n,k} p_{i,n,k} + \sum_{i \in T} \sum_{k=1}^{K} a_{i,k} \left( b_{i,k} \frac{D_i}{\chi_i} - R_{i,k} \right) + \sum_{i \in T} \sum_{k=1}^{K} \sum_{n=1}^{N} w_{i,n,k} p_{i,n,k} - P^w$$

(1-9)

Where $\alpha = \{\alpha_{i,k}, i \in T', k \in \kappa\}$ and $\beta = \{\beta_{i}, i \in T'\}$ are Lagrange multiplier factors. It should be noted that $W$ is set to a binary variable matrix and the value of the transmission power allocation strategy $P$ is set to non-negative, so the constraint condition $C5$ is automatically satisfied, and $C5, C6$ is ignored when constructing the Lagrangian function, of which $C6$. It indicates that one subcarrier is allocated to only one user.

For a given offload strategy $b$, the optimal subcarrier and its power allocation strategy can be obtained by solving problem $\rho^{2-1}$. The following conditions are sufficient and necessary conditions for subcarrier power optimization:

$$\frac{\partial \mathcal{R}(\alpha, \beta, W, P)}{\partial p_{i,n,k}} \bigg|_{p_{i,n,k}=P^*_n} = 0$$

(1-10)

After the partial derivative of function $\mathcal{R}(\alpha, \beta, W, P)$ to $p_{i,n,k}$, the above equation is converted to:

$$b_{i,k} \chi_i w_{i,n,k} - \frac{\alpha_{i,k} B_{i} w_{i,n,k} g_{i,n,k}}{\ln 2(\delta^2 + g_{i,n,k} p_{i,n,k})} + \beta_{i} w_{i,n,k} = 0$$

(1-11)

When user $i$ offloads tasks to MS $k$, the optimal power allocated to user subcarrier $n$ is:

$$p_{i,n,k}^* = \left[ \frac{\alpha_{i,k} B_{i} w_{i,n,k}}{\ln 2(b_{i,k} \chi_i + \beta_{i})} - \frac{\delta^2}{g_{i,n,k}} \right]^+$$

(1-12)
According to the obtained optimal subcarrier power allocation strategy \( p^∗(b) \), the partial derivative of \( W_{i,n,k}^∗ \) can be obtained as follows:

\[
\frac{\partial \mathcal{R}(\alpha, \beta, W, p^∗(b))}{\partial W_{i,n,k}}\bigg|_{n_{i,n,k}} = (b_{i,k}x_i + \beta_i)p^∗_{i,n,k} - \alpha_{i,k}B_n \log_2(1 + \frac{g_{i,n,k}^∗p^∗_{i,n,k}}{\delta^2})
\]

(1-13)

Since the partial derivative of \( \mathcal{R}(\alpha, \beta, W, p^∗(b)) \) with respect to \( W_{i,n,k} \) is independent of \( W_{i,n,k}^∗ \), the optimal \( W_{i,n,k}^∗ \) cannot be obtained according to the above formula, so the constructor \( \phi_{i,n,k}^∗ \) is:

\[
\phi_{i,n,k}^∗ = (b_{i,k}x_i + \beta_i)p^∗_{i,n,k} - \alpha_{i,k}B_n \log_2(1 + \frac{g_{i,n,k}^∗p^∗_{i,n,k}}{\delta^2})
\]

(1-14)

According to the internal relationship between the monotonicity of the function and the derivative, \( \phi_{i,n,k}^∗ \) and \( W_{i,n,k}^∗ \) are easy to know. There are the following associations:

\[
\phi_{i,n,k}^∗ = \begin{cases} 
> 0, \text{ when } \phi_{i,n,k} = 0 \\
= 0, \text{ when } \phi_{i,n,k}^∗ \in (0,1) \\
< 0, \text{ when } \phi_{i,n,k}^∗ = 1
\end{cases}
\]

(1-15)

It can be seen from the above formula that the optimal \( W_{i,n,k}^∗ \) appears either at the edge of the feasible domain or in the middle of the feasible domain, and then a subcarrier is allocated to only one user according to the requirements of constraint condition \( C6 \). The optimal subcarrier allocation strategy \( W^∗(b) \) can pass the following rules get:

\[
W_{i,n,k}^∗ = \begin{cases} 
1, \text{ when } \phi_{i,n,k} = \min(\phi(b)) \\
0, \text{ other situations}
\end{cases}
\]

(1-16)

Where \( \phi(b) = \{\phi_{i,n,k}, i \in T^i\}, n \in \eta \) is a row vector containing all partial derivatives of the Langerangian function \( \phi_{i,n,k} \) for subcarrier \( n \). Lagrange multiplier factors \( \alpha \) and \( \beta \) are updated with corresponding sub-gradients.

\[
a_{i,k}(m+1) = a_{i,k}(m) + \mu_{a} \frac{\partial \mathcal{R}(\alpha, \beta, W, P)}{\partial a_{i,k}} = a_{i,k}(m) + \mu_{a} \left( b_{i,k} \frac{D}{x_i} - R_{i,k}\right)
\]

(1-17)
Among them $\mu_\alpha$ and $\nu_\beta$ are the update steps of $\alpha$ and $\beta$.

5.2. Unloading strategy optimization

Given the sub-carriers and their power allocation strategies $W^*(b)$ and $p^*(b)$ obtained through the optimization of the low-level problem, the upper-level problem can be written in the form of $\rho^3$:

$$\rho^3 : \min_{b} F(b, W^*(b), p^*(b)) = \sum_{i \in \mathcal{T}} \sum_{k=1}^{K} b_{i,k} E_{i,k}^r$$

s.t. $C1 - C3, C8$ \hspace{1cm} (1-19)

Obviously $\rho^3$ is a convex optimization problem. When the uninstall strategy meets the constraints $C1 - C3$, search for the feasible domain of uninstall strategy $b$ with the minimum system energy consumption as the goal to obtain the optimal uninstall strategy:

$$b^* = \arg\min_{b} F(b, W^*(b), p^*(b)) = \sum_{i \in \mathcal{T}} \sum_{k=1}^{K} b_{i,k} E_{i,k}^r$$ \hspace{1cm} (1-20)

6. Conclusion

This paper discusses the joint optimization of offload strategy and wireless resources in a multi-user multi-MS system. In a resource-constrained system, an energy-efficient offload strategy and a wireless resource joint optimization algorithm (EEJS) are proposed to achieve a balance between user diversity and MS diversity.

Acknowledgments

State Grid Shanxi Electric Power Company Science and technology projects(520531200004).

References

[1] Colson B, Marcotte P, Saved G. An overview of of bilbil optimization[J]. Annals of Operrates Research, 2007, 153(1): 235-256.

[2] Kumar K, Lu Y H. Cloud coomporting for for mobile usables: Canofloading computation at save save Energy?[J]. Computer, 2010 (4): 51-56.