IJE TRANSACTIONS B: Applications Vol. 32, No. 11, (November 2019) 1627-1633

International Journal of Engineering

Journal Homepage: www.ije.ir

Joint Allocation of Computational and Communication Resources to Improve Energy Efficiency in Cellular Networks

N. Nouri*, A. Entezari

Cell Lab, Department of Electrical Engineering, Yazd University, Yazd, Iran

PAPER INFO

Paper history:
Received 22 October 2018
Received in revised form 30 April 2019
Accepted 03 May 2019

Keywords:
Mobile Cloud Computing
Heteregeneous Network
Non-convex Function
Bandwidth Allocation
Convex Approximation

ABSTRACT

Mobile cloud computing (MCC) is a new technology that has been developed to overcome the restrictions of smart mobile devices (e.g. battery, processing power, storage capacity, etc.) to send a part of the program (with complex computing) to the cloud server (CS). In this paper, we study a multi-cell with multi-input and multi-output (MIMO) system in which the cell-interior users request service for their processing from a common CS. Also, the problem of the optimum offloading is considered as an optimization problem with optimization parameters including communication resources (such as bandwidth, transmit power and backhaul link capacity) and computational resources (such as the capacity of cloud server) in the downlink network. The main goal is to minimize the total energy consumption by mobile users (MUs) for processing with the delay limitation for each use. This issue leads to a non-convex problem and to solve the problem, we use successive convex approximation (SCA) method. We finally show that the joint optimization of these parameters leads to reducing the energy consumption of the network with simulation examples.

***doi**: 10.5829/ije.2019.32.11b.14*

1. INTRODUCTION[1]

With the increment of technology in mobile devices, popular applications are daily offered to network users which will be more complex and demand heavy computation. Despite enhancing technology in mobile devices and their applications, there are some challenges in their resources such as storage capacity, battery lifetime and computational capacity which restricts the application's usage. Recently, mobile cloud computing (MCC) has been suggested as an efficient solution to overcome the restriction in mobile devices to benefit from cloud computing (CC) potential in mobile computing (MC) [1-4]. It can be said that MCC is a combination of CC and MC [5]. Employing this method, we can send a part of the program which has complicated computing and difficult calculations, to the cloud server (CS) [6]. The advantage of employing this method is to diminish the amount of energy consumption by mobile users (MUs), which improve the battery lifetime and computing speed [7, 8]. Moreover, using this type of processing, MUs do not require to upgrade their mobile devices in terms of hardware and software.

Barbarossa et al. [9] studied a technique for the joint allocation of communication and computation resources in the single-user mode. Besides, the optimal resources allocation in the network is generalized as multi-user form by Barbarossa et al. [9]. Unlike the consideration of centralized structure for CS in literature [9, 10]; Barbarossa et al. [11], Chen [12] consider that the CS has a decentralized structure and they solve the problem of optimal resources allocation via game theory methods.

Nouri et al. [13] proposed an offloading framework which reduces the total cost of the network and formulated the task offloading problem as a joint optimization of the computational and communicational resources. In contrast, Sardellitti et al. [14] tried to assign the optimal bandwidth to MUs who request services from the CS, as well as computational resources. Furthermore, MUs can perform a part of their computing on their devices. After modeling the system in the form of an optimization problem, we observe that the problem has a

*Corresponding Author Email: *nimanouri68@gmail.com* (N. Nouri)

Please cite this article as: N. Nouri, A. Entezari, Joint Allocation of Computational and Communication Resources to Improve Energy Efficiency in Cellular Networks, International Journal of Engineering (IJE), IJE TRANSACTIONS B: Applications Vol. 32, No. 11, (November 2019) 1627-1633

non-convex form. Considering a logical and simplistic assumption for solving the problem, we employ a method called successive convex approximation (SCA) in this paper. Extensive simulation studies demonstrate the energy efficiency improvement of our proposed method and the superior performance over several existed schemes.

This paper is organized as follows. In section 2, we provide our system model and formulate the problem of the optimal resources allocation in mathematical form. In section 3, we discuss the proposed algorithm and solve the problem. In section 4, we provide the results using some simulation examples and finally, we conclude the paper in section 5.

2. SYSTEM MODEL

Figure 1. shows the proposed model which is considered a multi-cell network with multi-input multi-output (MIMO) including N_c cell. Moreover, there are M MUs and a base station (BS) in each cell where all BSs are connected to a common server with limited resources. Note that these servers provide computing and storage resources to MUs which is called CSes.

We assume that the backhaul capacity between BS and CS is limited. In addition, MUs in each cell have orthogonal spectral resources, i.e., there is no intra-cell interference between MUs in each cell. However, the effect of the inter-cell interference is considered between MUs of cells with different spectral resources.

Also, we indicate all MUs with the set of $\mathrm{N} \triangleq \{m_n : m = 1,\ldots,M\ ,\ n = 1,\ldots,N_c\}$ in which m_n denotes the MU which uses the spectral resource m in cell n. We also consider the number of the transmitted and received antennas as $N_{T_{m_n}}$ and N_{R_n} , respectively.

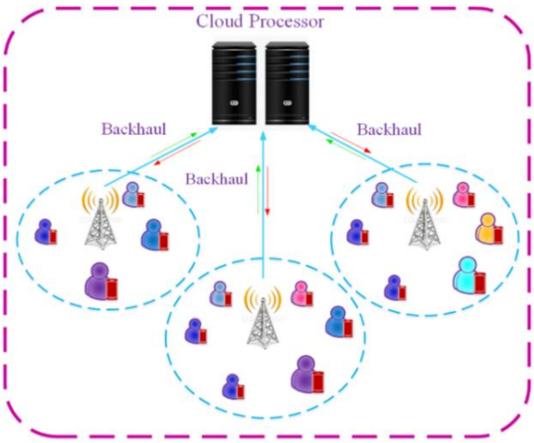

Figure 1. Proposed system model

We denote the application that each MU wants to run as $\mathrm{APP}_{m_n} = \{\mathrm{V}_{m_n}, \mathrm{B}^{I}_{m_n}, \tau^{\max}_{m_n}\}$ where V_{m_n} is required CPU-cycle to run the program. In addition, $\mathrm{B}^{I}_{m_n}$ and $\tau^{\max}_{m_n}$ indicate data bits (which includes transmitted code and additional data) and the upper bound of acceptable delay for running the application of each MU, respectively. We define the parameter λ_{m_n} as the processing percentage of each MU that transmits to the cloud server.

Therefore, the total delay of each MU will incur for receiving the service is given by:

$$\tau_{m_n}(\bullet) = \left(1-\lambda_{m_n}\right)\tau^{\ell}_{m_n}(\bullet) + \lambda_{m_n}\tau^{off}_{m_n}(\bullet), \tag{1}$$

where $\tau^{\ell}_{m_n}(\bullet)$ is the amount of caused delay to process the program local condition. Furthermore, $\tau^{off}_{m_n}(\bullet)$ in (1) denotes the total caused delay for receiving the service from the CS which can be expressed as follows:

$$\tau_{m_n} \triangleq \tau^{ul}_{m_n} + \tau^{bh}_{m_n} + \tau^{exe}_{m_n} + \tau^{dl}_{m_n}, \tag{2}$$

where $\tau^{ul}_{m_n}$ indicates the value of caused delay in transmitting data from MU m_n to the BS. $\tau^{exe}_{m_n}(\bullet)$ is the delay value that is consumed for computing the program in the CS and $\tau^{bh}_{m_n}(\bullet)$ is the caused delay in the backhaul between BS and CS in the downlink direction. Finally, $\tau^{dl}_{m_n}(\bullet)$ denotes the delay value for sending the transmitted processing results from the CS to the typical MU. Furthermore, energy consumption by each MU to receive the service is given by:

$$e_{m_n}(\bullet) = \lambda_{m_n}\left(e^{ul}_{m_n}(\bullet) + e^{dl}_{m_n}(\bullet)\right) + \left(1-\lambda_{m_n}\right)e^{\ell}_{m_n}(\bullet), \tag{3}$$

where $e^{ul}_{m_n}(\bullet)$ and $e^{dl}_{m_n}(\bullet)$ denote the energy that is transmitted and received data between typical MU and BS by MU m_n in the uplink and downlink directions, respectively. In addition, $e^{\ell}_{m_n}(\bullet)$ is the energy consumption in the CS condition.

The main purpose of this model is to minimize the total amount of energy consumption by MUs to receive the service with the delay limit constraint. In the sequel, we compute the values of the energy and delay and express the model in mathematical form.

2. 1. Local Processing

If we denote the computational capability of each MU in terms of the CPU-cycle per second by $f^{\ell}_{m_n}$, the required time for local computing APP_{m_n} in each MU can be derived as follows:

$$\tau^{\ell}_{m_n}(f^{\ell}_{m_n}) = \frac{V_{m_n}}{f^{\ell}_{m_n}}, m_n \in N, \tag{4}$$

In addition, the required energy for computing can be expressed as:

$$e^{\ell}_{m_n}(f^{\ell}_{m_n}) = \kappa V_{m_n}(f^{\ell}_{m_n})^2, m_n \in N, \tag{5}$$

in which κ is the effective capacitance of switch which depends on the structure of each MU [15].

2. 2. Uplink Transmission We assume that the transmitted signal from each MU is denoted by $\mathbf{X}^{ul}_{m_n}$ where $\mathbf{X}^{ul}_{m_n} \sim CN\left(0, Q^{ul}_{m_n}\right)$ and $Q^{ul}_{m_n} = \mathbf{E}\left[\mathbf{X}^{ul}_{m_n}\mathbf{X}^{ul\,\mathrm{H}}_{m_n}\right]$. Moreover, we express the feasible set of all covariance matrices $\mathbf{Q}^{ul}_{m_n}$ as follows:

$$\mathbf{Q}^{ul}_{m_n} \triangleq \left\{Q^{ul}_{m_n} \in C^{N_{T_{m_n}} \times N_{T_{m_n}}} : Q^{ul}_{m_n} \geq \mathbf{0}, tr(Q^{ul}_{m_n}) \leq P^{ul}_{m_n}\right\} \tag{6}$$

where $P^{ul}_{m_n}$ expresses the maximum power of each MU in the uplink direction. The data transmission rate of the MU m_n in terms of bits/seconds is given by:

$$r^{ul}_{m_n} = w^{ul}_{m_n} \log_2 \det(\mathbf{I} + \mathbf{H}^{\mathrm{H}}_{m_n n}\mathbf{R}^{ul}_{m_n}(\mathbf{Q}^{ul}_{-m_n}, w^{ul}_{m_n})\mathbf{H}_{m_n n} Q^{ul}_{m_n}) \tag{7}$$

where

$$\mathbf{R}^{ul}_{m_n}(\mathbf{Q}^{ul}_{-m_n}, w^{ul}_{m_n}) \triangleq w^{ul}_{m_n}\mathbf{N_0}\mathbf{I} + \sum_{j_r \in N, r \neq n} \mathbf{H}_{j_r,n} Q^{ul}_{j_r}\mathbf{H}^{\mathrm{H}}_{j_r,n} \tag{8}$$

in which $\mathbf{R}^{ul}_{m_n}(\mathbf{Q}^{ul}_{-m_n}, w^{ul}_{m_n})$ is the covariance matrix of the disturbance (noise plus inter-cell interference) in cell n and m^{th} spectral resource. In addition, $\mathbf{H}_{m_n,n}$ is the channel matrix between MU m_n and the tagged BS while $\mathbf{H}_{j_r,n}$ is the channel matrix between the interference MU j_r and the BS in cell n in the uplink case. $\mathbf{N_0}$ denotes the power spectral density of the noise and $w^{ul}_{m_n}$ denotes the bandwidth that is allocated to the MU m_n in the uplink case. We also have:

$$\mathbf{Q}^{ul}_{-m_n} \triangleq \left(\left(Q^{ul}_{j_r}\right)_{j=1}^{M}\right)_{r=1, r\neq n}^{N_c}. \tag{9}$$

The required time for transmitting $B^{I}_{m_n}$ data bits from MU to the BS can be expressed as:

$$\begin{aligned}\tau^{ul}_{m_n}(Q^{ul}_{m_n}, \mathbf{Q}^{ul}_{-m_n}, w^{ul}_{m_n}) &= \frac{B^{I}_{m_n}}{r^{ul}_{m_n}} \\ &= \frac{B^{I}_{m_n}}{w^{ul}_{m_n}\log_2\det\left(\mathbf{I} + \mathbf{H}^{\mathrm{H}}_{m_n,n}\mathbf{R}^{ul}_{m_n}(\mathbf{Q}^{ul}_{-m_n}, w^{ul}_{m_n})\mathbf{H}_{m_n,n}Q^{ul}_{m_n}\right)}.\end{aligned} \tag{10}$$

The energy consumption of the MU for transmitting data in the uplink case is given by:

$$\begin{aligned}e^{ul}_{m_n}(Q^{ul}_{m_n}, \mathbf{Q}^{ul}_{-m_n}, w^{ul}_{m_n}) &= tr\left(Q^{ul}_{m_n}\right)\tau^{ul}_{m_n}(Q^{ul}_{m_n}, \mathbf{Q}^{ul}_{-m_n}, w^{ul}_{m_n}) \\ &= \frac{B^{I}_{m_n} tr\left(Q^{ul}_{m_n}\right)}{w^{ul}_{m_n}\log_2\det\left(\mathbf{I} + \mathbf{H}^{\mathrm{H}}_{m_n,n}\mathbf{R}^{ul}_{m_n}(\mathbf{Q}^{ul}_{-m_n}, w^{ul}_{m_n})\mathbf{H}_{m_n,n}Q^{ul}_{m_n}\right)}.\end{aligned} \tag{11}$$

2. 3. Computing In CS We assume that the value of the CS computation capacity in terms of the CPU-cycle per second is equal to F^{Cloud}. Furthermore, $f^{C}_{m_n} \geq 0$ indicates the percentage of the total CS capacity which is assigned to the MU m_n, then $\sum_{m_n \in N} f^{C}_{m_n} \leq 1$. Therefore, the required duration to run the CPU-cycle for MU m_n is given by:

$$\tau^{ul}_{m_n}\left(f^{C}_{m_n}\right) = \frac{V_{m_n}}{f^{C}_{m_n}F^{Cloud}}. \tag{12}$$

2. 4. Backhaul Link Transmission We consider that the backhaul link capacity between the BS and CS is limited and the value of the capacity in terms of bits per second is denoted by C^{ul}_{n}. Moreover, $c^{ul}_{m_n} \geq 0$ is the percentage of these resources that are allocated to the MU m_n in the uplink direction, so $\sum_{m_n \in N} c^{ul}_{m_n} \leq 1$. Therefore, the delay value of each MU will incur on a Backhaul link can be calculated as :

$$\tau^{bh}_{m_n}\left(c^{ul}_{m_n}\right) = \frac{B^{I}_{m_n}}{c^{ul}_{m_n}C^{ul}_{n}}. \tag{13}$$

2. 5. Downlink Transmission Note that the delay and energy consumption of outcome from the CS to the MU are neglected in this model since the size of the outcome details is much smaller than the size of the input data $\left(B^{O}_{m_n} \ll B^{I}_{m_n}\right)$ that is similar to much existing research.

2. 6. Problem Statement In Form Of Optimization Problem At first, for simplicity, we gathered the optimization variables in vector $\mathbf{S}$ as follows:

$$\mathbf{S} \triangleq \left(\mathbf{Q}^{\mathrm{ul}}, \mathbf{w}^{\mathrm{ul}}, \mathbf{f}^{\mathrm{local}}, \mathbf{f}^{\mathrm{cloud}}, \mathbf{c}^{\mathrm{ul}}, \boldsymbol{\lambda}\right), \tag{14}$$

where

$$\begin{aligned}&\mathbf{Q}^{\mathrm{ul}} \triangleq \left(Q^{ul}_{m_n}\right)_{m_n \in N}, \mathbf{w}^{\mathrm{ul}} \triangleq \left(w^{ul}_{m_n}\right)_{m_n \in N} \\ &\mathbf{f}^{\mathrm{local}} \triangleq \left(f^{\ell}_{m_n}\right)_{m_n \in N}, \mathbf{f}^{\mathrm{cloud}} \triangleq \left(f^{C}_{m_n}\right)_{m_n \in N} \\ &\mathbf{c}^{\mathrm{ul}} \triangleq \left(c^{ul}_{m_n}\right)_{m_n \in N}, \boldsymbol{\lambda} \triangleq \left(\lambda_{m_n}\right)_{m_n \in N}.\end{aligned} \tag{15}$$

The optimal offloading problem can be expressed as an optimization problem in the form of minimizing the total energy consumption by all MUs with the delay constraint as follows:

$$\min_{\mathbf{Q}^{\mathrm{ul}},\mathbf{w}^{\mathrm{ul}},\mathbf{f}^{\mathrm{local}},\mathbf{f}^{\mathrm{cloud}},\mathbf{c}^{\mathrm{ul}},\lambda} \mathbf{E}^{\mathrm{tot}}(\mathbf{Q}^{\mathrm{ul}},\mathbf{w}^{\mathrm{ul}},\mathbf{f}^{\mathrm{local}}) = \sum_{m_n\in\mathrm{N}}\left(\lambda_{m_n}e^{ul}_{m_n}(\mathbf{Q}^{\mathrm{ul}},\mathbf{w}^{\mathrm{ul}})+\left(1-\lambda_{m_n}\right)e^{\ell}_{m_n}(f^{\ell}_{m_n})\right)$$

s.t.

C1. $\left(1-\lambda_{m_n}\right)\tau^{\ell}_{m_n}+\lambda_{m_n}\tau^{off}_{m_n}\le\tau^{\max}_{m_n},\forall m_n\in\mathrm{N}$

C2. $\sum_{m_n\in\mathrm{N}} w^{ul}_{m_n}\le\mathbf{W}^{ul},w^{ul}_{m_n}\ge 0,\forall m_n\in\mathrm{N}$

C3. $\sum_{m_n\in\mathrm{N}} f^{C}_{m_n}\le 1, f^{C}_{m_n}\ge 0,\forall m_n\in\mathrm{N}$

C4. $0\le f^{\ell}_{m_n}\le f^{\max}_{m_n},\forall m_n\in\mathrm{N}$

C5. $\sum_{m_n\in\mathrm{N}} c^{ul}_{m_n}\le 1, c^{ul}_{m_n}\ge 0,\forall m_n\in\mathrm{N}$

C6. $\mathrm{Q}^{ul}_{m_n}\in\mathbf{Q}^{ul}_{m_n},\forall m_n\in\mathrm{N}$

C7. $\lambda_{m_n}\in[0,1],\forall m_n\in\mathrm{N}$ $\qquad$ P 1

Where $C1$ denotes the delay value for receiving the service for MU in which should be less than the upper bound of the acceptable delay for each MU. In addition, **C2** indicates the restriction of the network bandwidth. **C3** and **C4** denote the computation resources limitation of the CS and local condition, respectively. Moreover, **C5** is the limitation of the backhaul link between BS and CS in the uplink direction. The problem P 1 is a non-convex optimization problem and the reason for the non-convexity of the problem is the fact that the objective function and **C1** are not convex. In the following, we evaluate the non-convex problem using SCA method.

3. PROBLEM SOLVING VIA SCA METHOD

with regards to the non-convexity of the objective function and **C1**, the problem P 1 is non-convex. Therefore, we use the SCA scheme [16] to solve the optimization problem. In this method, to derive a result, we employ an iterative algorithm that obtains a convex approximation for the non-convex expression in each iteration. It is worth to mention that the obtained approximations should satisfy the mentioned conditions in [16]. We next derive a convex approximation for the objective function and constraint **C1** so that satisfy the conditions where mentioned in [16].

3. 1. Convex Approximation Of The Objective Function

We consider the feasible set $\mathbf{K}$ such that all functions in P 1 are well defined on it. It must be noted that such a set always exists. If we denote the convex approximation of the objective function $\mathbf{E}^{\mathrm{tot}}(\mathbf{Q}^{\mathrm{ul}},\mathbf{w}^{\mathrm{ul}},\mathbf{f}^{\mathrm{local}})$ around the point $\mathbf{S}(\mathrm{V})$ as $\tilde{\mathbf{E}}^{\mathrm{tot}}\left(\mathbf{S},\mathbf{S}(\mathrm{V})\right)$, the approximation is obtained as:

$$\tilde{\mathbf{E}}^{\mathrm{tot}}\left(\mathbf{S},\mathbf{S}(\mathrm{V})\right)=\bar{\mathbf{E}}^{\mathrm{tot}}\left(\mathbf{S},\mathbf{S}(\mathrm{V})\right)+\sum_{m_n\in\mathrm{N}}\tilde{\mathbf{E}}_{m_n}\left(\mathbf{Q}^{ul}_{m_n},\mathbf{Q}^{ul}_{-m_n},w^{ul}_{m_n},f_{m_n};\mathrm{S}(\mathrm{V})\right)\cdot \tag{16}$$

where

$$\begin{aligned}&\tilde{\mathbf{E}}_{m_n}\left(\mathbf{Q}^{ul}_{m_n},\mathbf{Q}^{ul}_{-m_n},w^{ul}_{m_n},f_{m_n};\mathrm{S}(\mathrm{V})\right)=\ldots\\&\kappa\left(1-\lambda_{m_n}(\mathrm{V})\right)\mathrm{V}_{m_n}(f^{\ell}_{m_n})^2+\kappa(1-\lambda_{m_n})\mathrm{V}_{m_n}\left(f^{\ell}_{m_n}(\mathrm{V})\right)^2\\&+\frac{\mathrm{B}^{I}_{m_n}\lambda_{m_n}tr\left(\mathrm{Q}^{ul}_{m_n}(\mathrm{V})\right)}{w^{ul}_{m_n}(\mathrm{V})\log_2\det\left(\mathbf{I}+\mathbf{H}^{\mathrm{H}}_{m_n,n}\mathbf{R}^{ul}_{m_n}(\mathbf{Q}^{ul}_{-m_n}(\mathrm{V}),w^{ul}_{m_n}(\mathrm{V}))\mathbf{H}_{m_n,n}\mathrm{Q}^{ul}_{m_n}(\mathrm{V})\right)}\\&+\frac{\mathrm{B}^{I}_{m_n}\lambda_{m_n}(\mathrm{V})tr\left(\mathrm{Q}^{ul}_{m_n}\right)}{w^{ul}_{m_n}(\mathrm{V})\log_2\det\left(\mathbf{I}+\mathbf{H}^{\mathrm{H}}_{m_n,n}\mathbf{R}^{ul}_{m_n}(\mathbf{Q}^{ul}_{-m_n}(\mathrm{V}),w^{ul}_{m_n}(\mathrm{V}))\mathbf{H}_{m_n,n}\mathrm{Q}^{ul}_{m_n}(\mathrm{V})\right)}\\&+\frac{\mathrm{B}^{I}_{m_n}\lambda_{m_n}(\mathrm{V})tr\left(\mathrm{Q}^{ul}_{m_n}(\mathrm{V})\right)}{w^{ul}_{m_n}\log_2\det\left(\mathbf{I}+\mathbf{H}^{\mathrm{H}}_{m_n,n}\mathbf{R}^{ul}_{m_n}(\mathbf{Q}^{ul}_{-m_n}(\mathrm{V}),w^{ul}_{m_n}(\mathrm{V}))\mathbf{H}_{m_n,n}\mathrm{Q}^{ul}_{m_n}(\mathrm{V})\right)}\\&+\frac{\mathrm{B}^{I}_{m_n}\lambda_{m_n}(\mathrm{V})tr\left(\mathrm{Q}^{ul}_{m_n}(\mathrm{V})\right)}{w^{ul}_{m_n}(\mathrm{V})\log_2\det\left(\mathbf{I}+\mathbf{H}^{\mathrm{H}}_{m_n,n}\mathbf{R}^{ul}_{m_n}(\mathbf{Q}^{ul}_{-m_n}(\mathrm{V}),w^{ul}_{m_n})\mathbf{H}_{m_n,n}\mathrm{Q}^{ul}_{m_n}(\mathrm{V})\right)}\\&+\frac{\mathrm{B}^{I}_{m_n}\lambda_{m_n}(\mathrm{V})tr\left(\mathrm{Q}^{ul}_{m_n}(\mathrm{V})\right)}{w^{ul}_{m_n}(\mathrm{V})\log_2\det\left(\mathbf{I}+\mathbf{H}^{\mathrm{H}}_{m_n,n}\mathbf{R}^{ul}_{m_n}(\mathbf{Q}^{ul}_{-m_n}(\mathrm{V}),w^{ul}_{m_n}(\mathrm{V}))\mathbf{H}_{m_n,n}\mathrm{Q}^{ul}_{m_n}\right)}\\&+\sum_{j_p\in\mathrm{N},p\neq n}\langle\nabla_{\mathbf{Q}^{ul}_{m_n}}\tilde{\mathbf{E}}_{j_p}(\bullet),\mathbf{Q}^{ul}_{m_n}-\mathbf{Q}^{ul}_{m_n}(\mathrm{V})\rangle, \qquad (17)\end{aligned} \tag{17}$$

and

$$\bar{\mathbf{E}}^{\mathbf{tot}}\left(\mathbf{S},\mathbf{S}(\mathrm{V})\right)=\left(\mathbf{S}-\mathbf{S}(\mathrm{V})\right)^{\mathrm{T}}\xi\left(\mathbf{S}-\mathbf{S}(\mathrm{V})\right), \tag{18}$$

where the matrix ξ is a diagonal matrix with non-negative elements that can be determined as:

$$\xi\triangleq diag(\varepsilon_{q^{ul}},\varepsilon_{w^{ul}},\varepsilon_{f^{\ell}},\varepsilon_{f^{\mathrm{cloud}}},\varepsilon_{c^{ul}},\varepsilon_{\lambda}). \tag{19}$$

in which $\langle\mathbf{A},\mathbf{B}\rangle\triangleq\mathrm{Re}\left\{\mathrm{tr}\left(\mathbf{A}^{\mathbf{H}}\mathbf{B}\right)\right\}$. In (16), the second expression of the right-hand side is used for convexification of the objective function and $\bar{\mathbf{E}}^{\mathrm{tot}}\left(\mathbf{S},\mathbf{S}(\mathrm{V})\right)$ is added to make $\tilde{\mathbf{E}}_{m_n}$ strongly convex.

3. 2. Convex Approximation Of C1

In order to calculate the convex approximation of **C1**, we first rewrite it as follows:

$$\begin{aligned}&\lambda_{m_n}\tau^{\ell}_{m_n}+\left(1-\lambda_{m_n}\right)\tau^{off}_{m_n}=\\&\lambda_{m_n}\left(\tau^{ul}_{m_n}+\tau^{bh}_{m_n}+\tau^{exe}_{m_n}+\tau^{dl}_{m_n}\right)+\left(1-\lambda_{m_n}\right)\frac{\mathrm{V}_{m_n}}{f^{\ell}_{m_n}}\\&=\frac{\mathrm{B}^{I}_{m_n}\lambda_{m_n}}{r^{ul}_{m_n}}+\frac{\mathrm{B}^{I}_{m_n}\lambda_{m_n}}{c^{ul}_{m_n}C^{ul}_{n}}+\frac{\mathrm{V}_{m_n}\lambda_{m_n}}{f^{C}_{m_n}F^{Cloud}}+\left(1-\lambda_{m_n}\right)\frac{\mathrm{V}_{m_n}}{f^{\ell}_{m_n}}.\end{aligned} \tag{19}$$

Now, we define $\mathrm{J}(\bullet)\triangleq\frac{r^{ul}_{m_n}}{\lambda_{m_n}}$. If we indicate the first-order Taylor series approximation as $\bar{\mathrm{J}}(\bullet)$, we observe that

$\frac{\mathrm{B}_{m_n}^{I}}{\mathrm{J}^{-}(\bullet)}$ has a convex form. Therefore, regarding,

$$\frac{a}{b}=\frac{1}{2}\left(a+\frac{1}{b}\right)^{2}-\frac{1}{2}\left(a^{2}+\frac{1}{b^{2}}\right)\quad \forall a\geq 0, b>0 \tag{20}$$

the right side of this equality is the differential of two convex functions. Accordingly, with linearizing the concave part of (20), i.e., the left side of (20), we can obtain a locally tight convex upper bound as [16]:

$$\frac{a}{b}\leq\frac{1}{2}\left(a+\frac{1}{b}\right)^{2}-\frac{1}{2}\left(a^{\upsilon^{2}}+\frac{1}{b^{\upsilon^{2}}}\right)-a^{\upsilon}\left(a-a^{\upsilon}\right)+\frac{1}{b^{\upsilon^{3}}}\left(b-b^{\upsilon}\right). \tag{21}$$

By employing (21) in each term of (19), we can obtain the desired convex upper bound for (19).It can be easily seen that the evaluated approximations for the objective function and **C1** satisfies the conditions mentioned in [16]. Calculating these approximations and substituting them, the convex approximation of **C1** is derived and is denoted by $\tilde{\tau}_{m_n}(\bullet)$. Now, we are ready to solve the problem P1.

3. 2. Convex Approximation of Problem

Calculating the convex approximations of the objective function and **C1** around the feasible point $\mathrm{S}(\upsilon)$, we can solve the problem using SCA iterative algorithm instead of solving the problem P1.

$$\mathbf{S}^{\mathbf{opt}}=\min_{\mathbf{Q}^{ul},\mathbf{w}^{ul},\mathbf{f}^{local},\mathbf{f}^{cloud},\mathbf{c}^{ul},\boldsymbol{\lambda}}\tilde{\mathbf{E}}^{tot}\left(\mathbf{S},\mathbf{S}(\upsilon)\right)$$

$$s.t.$$

$$\textbf{C1.}\ \tilde{\tau}_{m_n}\left(\mathbf{S},\mathbf{S}(\upsilon)\right)\leq\tau_{m_n}^{\max},\forall m_n\in\mathrm{N}$$

$$\textbf{C2}\sim\textbf{C6}\ of\ \mathrm{P1} \qquad \mathrm{P2}$$

where $\mathbf{S}^{\mathbf{opt}}$ denote the final result of the problem. The SCA method is summarized in **Algorithm 1**.

Algorithm 1: SCA Solution for **P2**

Initialization: $\mathbf{S}(0)\in\mathbf{K}$;; $\gamma(\upsilon)\in(0,1]$; $\upsilon=0$,

1: If $\mathbf{S}(\upsilon)$ satisfies the termination criterion, stop.

2: Compute $\mathbf{S}(\upsilon)$ from P 2.

3: Set $\mathbf{S}(\upsilon+1)=\mathbf{S}(\upsilon)+\gamma(\upsilon)\left(\mathbf{S}^{opt}(\upsilon)-\mathbf{S}(\upsilon)\right)$.

4: Set $\upsilon\leftarrow\upsilon+1$, and return to step 1.

Output: $\mathbf{S}^{\mathbf{optimum}}=\left(\hat{\mathbf{Q}}^{ul},\hat{\mathbf{w}}^{ul},\hat{\mathbf{f}}^{local},\hat{\mathbf{f}}^{cloud},\hat{\mathbf{c}}^{ul},\hat{\boldsymbol{\lambda}}\right)$.

In this algorithm, $\mathbf{S}(0)$ is the initial point that is selected from the feasible region of the problem, i.e., $\mathbf{K}$. Also, $\mathbf{S}^{opt}(\upsilon)$ denotes the optimal result in iteration υ. The stopping criteria of the algorithm is $\left|\tilde{\mathbf{E}}^{tot}\left(\mathrm{S}(\upsilon+1)\right)-\tilde{\mathbf{E}}^{tot}\left(\mathrm{S}(\upsilon)\right)\right|\leq\delta$ in which δ determines the algorithm accuracy. Furthermore, γ determines the algorithm step where $\gamma(\upsilon)=\left(1-\alpha\gamma(\upsilon-1)\right)\gamma(\upsilon-1)$, $\gamma(0)\in(0,1]$ and $\alpha\in\left(0,\frac{1}{\gamma(0)}\right)$.

4. SIMULATION RESULTS

We consider a network with two cells that there are two MUs in each cell, i.e., $M=N_c=2$. We assume that the number of the transmitted and received antennas are two ($N_{R_n}=N_{T_{m_n}}=2$). The other simulation parameters are $\mathbf{W}^{ul}=10$ MHz, $C_n^{ul}=10$ Mbits/ s , $F^{Cloud}=10^{11}$ CPU-cycle per sec, $\mathrm{V}_{m_n}=2640\times\mathrm{B}_{m_n}^{I}$ CPU-cycle/sec and $\mathcal{B}_{i_n}^{I}$ has a uniform distribution in $(0.1,1]$ Mbits.

Figure 3. shows the value of the total energy consumption in the network according to algorithm iteration. As observed in Figure 3., when MUs use partial offloading to receive the service, the value of the energy consumption is lower than the other items.

Figure 4. shows the value of the total network energy consumption in terms of the upper bound of the acceptable MUs delay. As expected, the amount of energy consumption reduces with the increment of the delay upper bound. However, if the upper bound is very small, the value of energy consumption increases proportionately.

Figure 5. illustrates the total network energy consumption according to the upper bound of the acceptable MUs delay in three modes: local processing, cloud processing and joint processing (the combination of the cloud and local processing). As shown in Figure 5., by the joint allocation of resources in partial processing, the value of the total network energy consumption is significantly diminished compared to the local and cloud processing. For example, with $\tau_{m_n}^{\max}=0.15$ sec, the value of the energy consumption is reduced to about 65% and 35% compared to the local processing and cloud processing, respectively.

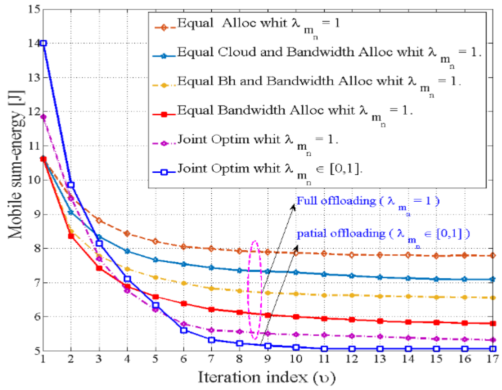

Figure 3. The total network energy consumption in terms of algorithm iteration

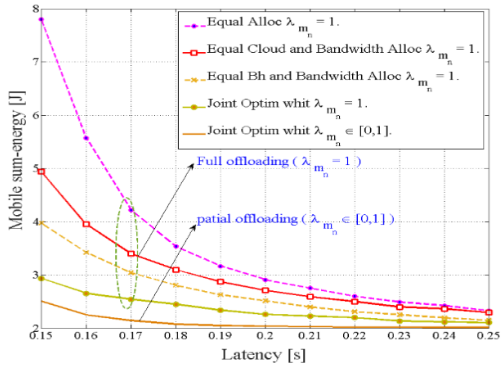

Figure 4. Energy consumption versus the upper bound of the acceptable delay.

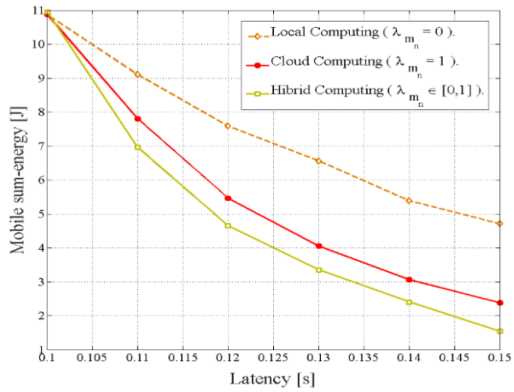

Figure 5. Comparison of total network energy consumption in the local, cloud and hybrid processing.

5. CONCLUSIONS

In this paper, we investigated the optimal allocation of the resources in a multi-cell network with MIMO. The assigned resources were communication and computational resources. The main goal of this model was to minimize the value of energy consumption with delay constraint. We expressed the problem of the resources optimal allocation in the form of the optimization problem in mathematical form. Since the problem was non-convex, we employed the SCA iterative algorithm for solving the problem. We assumed that the backhaul link capacity between BSs and CS was restricted and the MUs could send a part of the processing for computing. The simulation results showed that the jointly resources optimal allocation led to lower energy consumption. Also, the delay value was significantly reduced. Furthermore, the number of MUs was able to receive the service from CS, was increased.

6. REFERENCES

1. Jeyanthi, N., Shabeeb, H., Durai, M.S. and Thandeeswaran, R., "Rescue: Reputation-based service for cloud user environment", ***International Journal of Engineering-Transactions B: Applications***, Vol. 27, No. 8, (2014), 1179-1186.
2. Gubbi, J., Buyya, R., Marusic, S., Palaniswmi, M., "Internet of Things (IoT): A vision, architectural elements, and future directions", ***Future Generation Computer Systems***, Vol. 29, No. 7, (2013), 1645-1660.
3. Palacin, M., "Recent advances in rechargeable battery materials: a chemists perspective", ***Chemical Society Reviews***, Vol. 38, No. 9, (2009), 2565-2575.
4. Abolfazli, S., Sanaei, Z., Ahmed, E., Gani, A., Buyya, R., "Cloud-based augmentation for mobile devices: motivation, taxonomies, and open challenges", ***IEEE Communications Surveys & Tutorials***, Vol. 16, No. 1, (2014), 337-368.
5. Fernando, N., Loke, S., Rahayu, W., "Mobile cloud computing: A survey", ***Future Generation Computer Systems***, Vol. 29, No. 1, (2013), 84-106.
6. Liu, Yanchen, Myung, Lee., Zheng, Yanyan., "Adaptive multi-resource allocation for cloudlet-based mobile cloud computing system", ***IEEE Transactions on Mobile Computing***, Vol. 15, No. 10, (2016), 2398-2410.
7. Song, Weiguang, Xiaolong, Su., "Review of mobile cloud computing", ***Communication Software and Networks (ICCSN), 2011 IEEE 3rd International Conference on. IEE***, (2011), 1-4.
8. Satyanarayanan, M., et al, "The case for VM-based cloudlets in mobile computing", ***IEEE pervasive Computing***, (2009), 14-23.
9. Barbarossa, S., Sardellitti, S., Di Lorenzo, P., "Computation offloading for mobile cloud computing based on wide cross-layer optimization", **in *Proc. of Future Network and Mobile Summit (FuNeMS2013)***, (2013), 1-10.
10. Barbarossa, Sardellitti, Di Lorenzo, "Joint allocation of computation and communication resources in multiuser mobile cloud computing", in ***Proc. of IEEE Workshop* on Signal Processing Advances in Wireless Communications (SPAWC*2013)***, Darmstadt, Germany, (2013), 26-30.
11. Barbarossa, Sardellitti, Di Lorenzo, "Communicating while computing: Distributed mobile cloud computing over 5G heterogeneous networks", ***IEEE Signal Processing Magazine***, Vol. 31, No. 6, (2014), 45-55.
12. Chen, X., "Decentralized computation offloading game for mobile cloud computing", ***IEEE Transactions on Parallel and Distributed Systems***, 2014.
13. Nouri, N., Rafiee, P., Tadaion, A., "NOMA-Based Energy-Delay Trade-Off for Mobile Edge Computation Offloading in 5G Networks", in 9th International Symposium on Telecommunications (IST) (2018), 522-527.
14. Sardellitti, Stefania, Scutari, G., Barbarossa, S., "Joint optimization of radio and computational resources for multicell mobile-edge computing", ***IEEE Transactions on Signal and Information Processing over Networks***, No. 2, (2015), 89-103.
15. Zhang, W., Wen, Y., Guan, K., Dan, K., Luo, H., Wu, D., "Energy optimal mobile cloud computing under stochastic wireless channel", ***IEEE Transactions on Wireless Communications***, Vol. 12, No. 9, (2013), 4569–4581.
16. Song, Weiguang, Su, X., "Parallel and distributed methods for constrained nonconvex optimization-part I: Theory", ***IEEE Trans. Signal Processing***, Vol. 65, No. 8, (2017), 1929-1944.

Joint Allocation of Computational and Communication Resources to Improve Energy Efficiency in Cellular Networks

N. Nouri, A. Entezari

Cell Lab, Department of Electrical Engineering, Yazd University, Yazd, Iran

PAPER INFO

Paper history:
Received 22 October 2018
Received in revised form 30 April 2019
Accepted 03 May 2019

Keywords:
Mobile Cloud Computing
Heteregeneous Network
Non-convex Function
Bandwidth Allocation
Convex Approximation

چکیده

چکیده – پردازش ابری موبایل یک فن‌آوری نوظهور است که برای غلبه بر محدودیت‌های گوشی‌های موبایل (مانند باتری، توان پردازشی، ظرفیت ذخیره‌سازی و غیره)، باهدف ارسال بخشی از برنامه (شامل پردازش‌های پیچیده و محاسبات سنگین) به سمت سرور ابری مطرح‌شده است. در این مقاله، سیستم چند سلولی شامل چند ورودی و چند خروجی مطرح‌شده است که کاربران درون سلول‌ها برای انجام پردازش خود از سرور ابری مشترک، تقاضای سرویس می‌کنند. همچنین، مسأله تخلیه بهینه به‌صورت یک مسأله بهینه‌سازی مطرح می‌شود که پارامترهای بهینه‌سازی شامل منابع مخابراتی (مانند پهنای باند، توان ارسالی و ظرفیت لینک بکهال) و منابع محاسباتی (مانند ظرفیت محاسباتی سرور ابری) درون شبکه در مسیر فراسو هستند. هدف اصلی این است که با رعایت قید تأخیر برای هر کاربر، مجموع کل انرژی مصرف‌شده برای انجام پردازش توسط کاربران شبکه حداقل شود. طرح این مسأله منجر به یک مسأله نامحدب خواهد شد که برای حل آن از روشی بنام SCA استفاده می‌شود. نتایج حاصل از شبیه‌سازی نشان‌دهنده این است که بهینه‌سازی توأم این پارامترها، منجر به کاهش میزان مصرف انرژی در شبکه می‌شود.

***doi**: 10.5829/ije.2019.32.11b.14*