Energy-Efficient Resource Allocation for Heterogeneous Services in OFDMA Downlink Networks: Systematic Perspective

Quansheng Xu, Xi Li, Hong Ji, *Senior Member, IEEE*, and Xiaojiang Du, *Senior Member, IEEE*

***Abstract*—In the area of energy-efficient (EE) resource allocation, only limited work has been done on joint consideration of transmitter and receiver energy consumption. In this paper, we propose a novel EE resource allocation scheme for orthogonal frequency division multiple access (OFDMA) networks, where both transmitter energy consumption and receiver energy consumption are considered. In addition, different quality of service (QoS) requirements including minimum-rate guarantee service and best-effort service are taken into account. The time slot, sub-carrier (frequency) and power allocation policies are jointly considered to optimize system energy efficiency. With all these considerations, the EE resource allocation problem is formulated as a mixed combinatorial and non-convex optimization problem, which is extremely difficult to solve. To reduce the computational complexity, we tackle this problem with three steps. First, for given power allocation, we obtain the time-frequency resource unit (RU) allocation policy via binary quantum-behaved particle swarm optimization (BQPSO) algorithm. Second, under the assumption of known RU allocation, we transform the original optimization problem into an equivalent concave optimization problem and obtain the optimal power allocation policy through the Lagrange dual approach. Third, an iteration algorithm is developed to obtain the joint time-frequency-power resource allocation strategy. We validate the convergence and effectiveness of the proposed scheme by extensive simulations.**

***Index Terms*—Energy efficiency, resource allocation, heterogeneous service, OFDMA network, mixed combinatorial and non-convex optimization.**

I. Introduction

WITH the explosive growth of high-data-rate applications, more and more energy is consumed in wireless networks. Due to limited energy supply and the need of environmental-friendly transmission behaviors [1]–[5], energy-efficient (EE) wireless communication is drawing increasing attention. Several international research projects dedicated to EE wireless communication are being carried out, such as Green Radio, EARTH, OPERA-Net, eWIN, and so on [1].

This work is supported by the National Nature Science Foundation (NSF) under Grant 61271182, the NSF for Young Scientists of China under Grant 61302080, the Beijing Higher Education Young Elite Teacher Project under Grant YETP0432, and the US NSF under grant CNS-1065444.

Q. Xu, X. Li and H. Ji are with the Key Laboratory of Universal Wireless Communication, Ministry of Education, Beijing University of Posts and Telecommunications, Beijing 100876, China (e-mail:{quansheng,lixi,jihong}@bupt.edu.cn).

X. Du is with the Dept. of Computer and Information Sciences, Temple University, Philadelphia, PA 19122 USA (e-mail: dxj@ieee.org).

EE wireless communication includes many research areas, such as low-power circuit design, high-efficiency power amplifier and digital signal processing (DSP) technologies, EE resource management, EE network architecture and planning, adequate EE metric and energy consumption model, adaptive traffic pattern and load variation algorithm, as well as advanced cooling systems [1]–[3]. As an important aspect of EE resource management, EE resource allocation is very significant for enhancing energy efficiency performance [6]–[8].

Lots of EE resource allocation algorithms have been proposed to maximize energy efficiency for different fading channels, such as frequency-selective fading channel, flat fading channel, et al. [9]–[11]. And it has been shown that a unique global maximum energy efficiency exists and can be obtained by the proposed algorithms [8]–[11]. Besides, some efficient resource allocation algorithms have been proposed to optimize the tradeoff between spectrum efficiency (SE) and energy efficiency [12]–[14], bandwidth consumption and energy consumption [15], [16], as well as delay performance and energy consumption [17], [18]. However, almost all the aforementioned algorithms only optimize base station (BS) energy consumption and do not consider user equipment (UE) receiver circuit energy consumption that used to receive and process downlink traffic, which can significantly increase the UE receiver circuit energy consumption [19] and result in low energy efficiency. Therefore, the proposed algorithms are not energy efficiency from the system perspective.

Energy supply of UE is limited. Discontinuous reception (DRX) technology is always used to save the circuit energy consumption of UE, because circuit energy consumption increases with data transmission time [20], [21]. With the inspiration of DRX, some resource allocation algorithms have been proposed [7], [22]. In orthogonal frequency division multiple access (OFDMA) systems, traffic to one UE can be scheduled into fewer time slots to reduce the energy consumption, then the authors in [7] propose a green resource allocation algorithm to minimize the total receiving energy consumption of UEs. The authors in [22] propose a DRX-aware scheduling method where DRX parameters are used for scheduling, so as to reduce packet loss rate and UE energy consumption. However, these researches only optimize UE energy consumption, which imposes a strict restriction on resource allocation, causes services not able to use the most suitable resource, and may result in low energy efficiency.

Energy efficiency enhancement at the system level can be achieved only if energy consumption of the entire communication chains are considered [19]. However, if UE circuit energy consumption and BS transmission energy consumption are not comparable, it is unnecessary to jointly consider transmitter and receiver for designing EE resource allocation algorithm. In traditional macro cellular scenarios, when transmission distance is large, circuit energy consumption of UE receiver is always much lower than BS transmission energy consumption. However, in many short-range wireless communication systems (e.g., femtocell, wireless sensor networks (WSN)), the circuit energy consumption of receiver becomes comparable to or even exceeds the transmission energy consumption [3], [20], [23]. Therefore, when designing EE resource allocation algorithms for short-range communication situations, it is feasible and even necessary to jointly consider transmitter and receiver energy consumption.

So far, few works have jointly considered transmitter and receiver energy consumption when designing EE resource allocation algorithms. Authors in [19] propose a packet scheduling algorithm that can minimize both BS transmission and UE circuit energy consumption, while meeting service quality of service (QoS) requirement. However, this work has several weaknesses. First, with the objective of minimizing BS transmission energy and UE circuit energy consumption, it is not necessarily energy efficiency [8], [23]. Second, to decrease UE circuit energy consumption, at the beginning of each scheduling period only a fraction of time is allowed to transmit data, which induces low resource utilization efficiency. Third, energy efficiency can be further improved, for it does not consider BS circuit energy consumption. Moreover, [24] proposes an end-to-end EE resource allocation algorithm, however, only heuristic method is used to solve the formulation problem.

In this paper, we investigate the performance optimization of energy efficiency for downlink communications in OFDMA networks from a systematic perspective, where the BS transmission, BS circuit and UE circuit energy consumption are all taken into account. The resource allocation problem is formulated as a mixed combinatorial and non-convex optimization problem, where the time slot, sub-carrier (frequency) and power allocation policies are considered together to optimize energy efficiency. To reduce the computational complexity of the formulated problem, we tackle this problem with three steps. Step 1, for given power allocation, we obtain the time-frequency resource unit (RU) allocation policy via binary quantum-behaved particle swarm optimization (BQPSO) algorithm [25]; Step 2, under the assumption of known RU allocation, we transform the original optimization problem to an equivalent concave optimization problem and obtain the optimal power allocation policy through the Lagrange dual approach; Step 3, based on step 1 and step 2, an iteration algorithm is developed to obtain the time-frequency-power resource allocation strategy. We validate the convergence and effectiveness of the proposed scheme by extensive simulations. The distinct features of this paper are summarized as follows:

- Different from most existing works, we consider EE resource allocation from a systematic perspective. In the problem formulation, the BS transmission, BS circuit and UE energy consumption are jointly considered, which can achieve better performance of energy efficiency.
- Heterogeneous services including minimum-rate guarantee service and best-effort service are supported by our proposed resource allocation scheme, which is realistic, for heterogeneous services may simultaneously request system resource.
- Since the time slot, sub-carrier and power resource are jointly considered in our problem formulation, the proposed scheme can be regarded as a multi-dimensional resource allocation scheme. The more resource dimensions we consider, the harder it is to solve the formulated problem. In fact, only few works have been done in multi-dimensional resource allocation.

The remainder of the paper is organized as follows. Section II gives the system model and problem formulation. In Section III, the *Time-Frequency RU Allocation for Given Power Allocation* is discussed. In Section IV the *Power Allocation for Given RU Allocation* is presented. The *Time-Frequency-Power Resource Allocation* is developed in Section V. The performance analysis and discussions are given in Section VI. Finally, we conclude this paper in Section VII.

II. System Model and Problem Formulation

In this section, we introduce the system model and formulate the problem of EE resource allocation. To make the rest of this paper easy to follow, we list some frequently used notations in TABLE I.

TABLE I: Some Notations Used in This Paper

N	number of sub-carriers
M	number of time slots in each scheduling period
K	number of users
K_1	number of users with minimum-rate guarantee service
W	bandwidth of each sub-carrier
T	duration of one time slot
Ω_A	set of users with minimum-rate guarantee service
Ω_B	set of users with best-effort service
R_k^{min}	minimum rate threshold for user k, $k \in \Omega_A$
η_k	proportional-fairness factor for user k, $k \in \Omega_B$
R_{tot}	total transmitted data during one scheduling period
P_{tot}	total energy consumption
P_{max}	transmission power budget of BS
R_k	transmitted data of user k
$R_{n,m,k}$	transmitted data of user k on the RU (n, m)
$\gamma_{n,m,k}$	SNR of unit transmission power, i.e., CNR
$a_{n,m,k}$	Boolean variable indicating the RU allocation
$p_{n,m,k}$	transmission power of user k on RU (n, m)
$\mathbf{A}$	RU allocation policy with the element $a_{m,n,k}$
$\mathbf{P}$	power allocation strategy with the element $p_{m,n,k}$
$\mathbb{A}_k$	set composed of RUs that are allocated to user k
P_c	circuit power of BS
P_{r_k}	circuit power of UE k at receiving mode
P_{nr_k}	circuit power of UE k at non-receiving mode
$\mathbf{D}$	particle position in BQPSO algorithm
$U(\cdot,\cdot)$	fitness function in BQPSO algorithm
$\mathbf{M}_{be}(\cdot)$	mean best position of all particles
$\mathbf{B}_i^{be}(\cdot)$	best position of the i-th particle
$\mathbf{G}_{be}(\cdot)$	global best position of all particles
$\mathbf{L}_i(\cdot)$	local attractor for particle i

A. System Model

A single cell OFDMA network with K users and N sub-carriers is considered. Assume that these K users have heterogeneous service requirements and can be classified into two classes: users with minimum-rate guarantee service and users with best-effort service [26]. The corresponding sets of these two user classes are denoted as $\Omega_A = \{1, \cdots, K_1\}$ and $\Omega_B = \{K_1+1, \cdots, K\}$, respectively. Assume each sub-carrier has a bandwidth of W and can be modeled as Rayleigh block fading. We further assume that the channel state information (CSI) can be estimated perfectly. A RU represents one sub-carrier in one time slot with duration T, and one scheduling period contains M time slots. At each beginning of scheduling period, the BS is responsible for allocating all the $N \times M$ RUs and power resource among the K users.

B. Problem Formulation

The classical performance metric of energy efficiency "bits-per-Joule" [4], i.e., the number of delivered bits per consuming unit energy, is adopted in the paper. This means that energy efficiency is defined as the amount of system transmitted data R_{tot} divided by the total energy consumption P_{tot}.

The amount of system transmitted data R_{tot} during one scheduling period is given as

$$R_{tot} = \sum_{k=1}^{K} R_k, \tag{1}$$

where, R_k is the amount of transmitted data of user k during one scheduling period, which can be expressed as

$$R_k = \sum_{m=1}^{M} \sum_{n=1}^{N} TW \log_2(1 + a_{n,m,k}\gamma_{n,m,k}p_{n,m,k}), \tag{2}$$

where, $a_{n,m,k}$ is the RU allocation indicator. $a_{n,m,k} = 1$ denotes that RU (n, m) is allocated to user k; otherwise, $a_{n,m,k} = 0$. $\gamma_{n,m,k} = |h_{n,m,k}|^2/N_0W$ is the signal to noise ratio (SNR) of unit transmission power, i.e., channel gain to noise ratio (CNR). $h_{n,m,k}$ denotes the channel gain of user k on RU (n, m), and N_0 represents single-sided noise power spectral density. $p_{n,m,k} \geq 0$ denotes the transmission power of user k on RU (n, m).

The total energy consumption P_{tot} of transmitting R_{tot} bits information can be calculated as following. The total number of time slots where there are data for user k can formulated as $M_{r_k}(\mathbb{A}_k) = \sum_{m=1}^{M} f(\sum_{n=1}^{N} a_{n,m,k})$, where $\mathbb{A}_k$ is a set composed of RUs that are allocated to user k. If we know the $\mathbf{A}$ we can obtain the set $\mathbb{A}_k$ easily. $f(x)$ is an integer step function, where $f(x) = 0$ when $x = 0$, and $f(x) = 1$ when $x \in \{1, \cdots, N\}$. To simply the analysis, assuming there are only two work modes of UE in downlink transmission, i.e., receiving mode and non-receiving mode. The circuit power of UE k at receiving mode and non-receiving mode are P_{r_k} and P_{nr_k}, respectively. Assuming the circuit power of BS is always P_c. Thus the total energy consumption P_{tot} can be given as

$$\begin{aligned} P_{tot} =& T\Big[\sum_{k=1}^{K}\sum_{m=1}^{M}\sum_{n=1}^{N} p_{n,m,k} + \sum_{k=1}^{K} P_{r_k} M_{r_k}(\mathbb{A}_k) \\ &+ \sum_{k=1}^{K}(M - M_{r_k}(\mathbb{A}_k))P_{nr_k}) + P_c M\Big]. \end{aligned} \tag{3}$$

Then the resource allocation problem formulation from the systematic perspective can be given as:

$$\begin{aligned} &\max_{\mathbf{A},\mathbf{P}} \frac{R_{tot}(\mathbf{A},\mathbf{P})}{P_{tot}(\mathbf{A},\mathbf{P})} \\ \texttt{s.t.}\quad & C1: a_{n,m,k} \in \{0,1\}, \quad \forall n,m,k, \\ & C2: \sum_{k=1}^{K} a_{n,m,k} \leq 1, \quad \forall n,m, \\ & C3: \sum_{k=1}^{K}\sum_{n=1}^{N} p_{n,m,k} \leq P_{max}, \quad \forall m, \\ & C4: R_k \geq R_k^{min}, \quad \forall k \in \Omega_A, \\ & C5: \frac{R_k}{\sum_{k=K_1+1}^{K} R_k} = \eta_k, \quad \forall k \in \Omega_B, \end{aligned} \tag{4}$$

where, $\mathbf{A}$ with element $a_{m,n,k}$ and $\mathbf{P}$ with element $p_{m,n,k}$ are RU allocation policy and power allocation strategy, respectively. They both are $N \times M \times K$ matrix. The constraint (C) 1 and C2 are RU allocation constraints. C2 means that one RU can only be assigned to one user at most. The C3 is power allocation constraint which gives the maximum transmission power of BS, and P_{max} is transmission power threshold. The C4 is used to guarantee the minimum-rate of user k in Ω_A, and R_k^{min} is the minimum rate threshold. The C5 can ensure the fairness of user k in Ω_B, and η_k are proportional-fairness factors which are predetermined values.

The optimal resource allocation problem in (4) is a mixed combinatorial and non-convex optimization problem. The combinatorial nature comes from the RU allocation constraints C1 and C2. The non-convexity feature is caused by the proportional-fairness constraint C5 and the fractional form of the objective function. Furthermore, the UE receiver energy consumption is considered and formulated as $T(\sum_{k=1}^{K} P_{r_k} M_{r_k}(\mathbb{A}_k) + \sum_{k=1}^{K}(M - M_{r_k}(\mathbb{A}_k))P_{nr_k})$, which is non-differential for arguments $a_{m,n,k}$. Therefore, the resource allocation problem is very difficult to solve. In this paper, to solve the problem and obtain the resource allocation policies, we develop the following three algorithms: *Time-Frequency RU Allocation for Given Power Allocation*, *Power Allocation for Given RU Allocation*, and *Time-Frequency-Power Resource Allocation*.

III. Time-Frequency RU Allocation for Given Power Allocation

In this section, for given power allocation, we present a RU allocation algorithm, which is based on BQPSO. BQPSO is a novel simulated evolvement algorithm, which can effectively solve complicated combinatorial optimization problem with desirable performance of finding global optimal solution [25]. First, we introduce the BQPSO algorithm, and then we present the BQPSO based RU allocation algorithm.

The BQPSO algorithm has three important parts, i.e., particle position, fitness function and evolution equation. The position of each particle represents the possible solution of the optimization problem. In this paper, the position of each particle represents the possible RU allocation policy, i.e., decides how to assign $N \times M$ RUs to K users. Therefore, the

$N \times M$ RUs are regarded as $N \times M$ decision variables, and each decision variable with $\lceil \log_2 K \rceil$ bits, where $\lceil \cdot \rceil$ means rounding up the value. The particle position is defined as (5), which is a binary string with the length of $N \times M \times \lceil \log_2 K \rceil$. The bits $\mathbf{D}^1 = (d_{1,1,1}, \cdots, d_{1,1,\lceil \log_2 K \rceil})$ in $\mathbf{D}$ belong to the first decision variable, i.e., the RU $(1, 1)$.

$$\mathbf{D} = (d_{1,1,1}, \cdots, d_{1,1,\lceil \log_2 K \rceil}, \cdots, d_{N,M,\lceil \log_2 K \rceil}). \tag{5}$$

According to the position $\mathbf{D}$, we can get the RU allocation policy $\mathbf{A}$. For example, the $n \times M + m$th decision variable, i.e., the RU (n, M), should be allocated to user kk, $kk = d_{n,m,1} 2^{\lceil \log_2 K \rceil - 1} + d_{n,m,2} 2^{\lceil \log_2 K \rceil - 2} + \cdots + d_{n,m,\lceil \log_2 K \rceil} 2^0 + 1$. That means, $a_{n,m,k} = 0$ if $k \neq kk$ and $a_{n,m,k} = 1$ if $k = kk$.

The fitness function is used to evaluate the quality of the obtained solution, which is constructed by the original optimization problem. Assume the power allocation policy is given as $\mathbf{P}^t$. Then using the method of penalty function, the fitness function is given as

$$U(\mathbf{A}, \mathbf{P}^t) = F(\mathbf{A}, \mathbf{P}^t) - \alpha F_p(\mathbf{A}, \mathbf{P}^t), \tag{6}$$

where $F(\mathbf{A}, \mathbf{P}^t) = R_{tot}(\mathbf{A}, \mathbf{P}^t)/P_{tot}(\mathbf{A}, \mathbf{P}^t)$ is the objective function, α is penalty factor, and $F_p(\mathbf{A}, \mathbf{P}^t)$ represents penalty function that consists of constraints related to $a_{m,n,k}$. The particle position in BQPSO is a binary string, and each RU is regarded as a decision variable allocated to one user at most, hence C1 and C2 in (4) have been included. Then the penalty function can be written as:

$$\begin{aligned} F_p(\mathbf{A}, \mathbf{P}^t) = & \sum_{k=1}^{K_1} [\max(0, R_k^{min} - R_k)]^2 \\ & + \sum_{k=K_1+1}^{K} (\eta_k \sum_{k=K_1+1}^{K} R_k - R_k)^2, \end{aligned} \tag{7}$$

where $\max(\cdot, \cdot)$ returns a larger value of the two variables.

Directly describing the evolution equation of BQPSO may be difficult to understand. Hence, firstly the evolution equation of quantum-behaved particle swarm optimization (QPSO) is introduced. Assume there are I particles in search space. The evolution equation of particle i $(i = 1, \cdots, I)$ in the QPSO algorithm is given as following [25], [27]:

$$\begin{cases} \mathbf{D}_i(l+1) = \mathbf{L}_i(l) + \upsilon |\mathbf{M}_{be}(l) - \mathbf{D}_i(l)| \cdot \ln(\frac{1}{u}) & \text{if } r \geq 0.5 \\ \mathbf{D}_i(l+1) = \mathbf{L}_i(l) - \upsilon |\mathbf{M}_{be}(l) - \mathbf{D}_i(l)| \cdot \ln(\frac{1}{u}) & \text{if } r < 0.5, \end{cases} \tag{8}$$

where l denotes the iteration time, υ is the contraction-expansion coefficient which can be used to control algorithm convergence rate, u and r are both random variables between 0 and 1. The $\mathbf{M}_{be}(l)$ is the mean best position of all particles in the l-th iteration, which can be obtained by

$$\mathbf{M}_{be}(l) = \frac{1}{I} \sum_{i=1}^{I} \mathbf{B}_i^{be}(l), \tag{9}$$

where $\mathbf{B}_i^{be}(l)$ is the best position of the i-th particle in the l-th iteration. The $\mathbf{L}_i(l)$ in (8) is called local attractor for particle i in the l-th iteration, which can be given as

$$\mathbf{L}_i(l) = \theta \mathbf{B}_i^{be}(l) + (1 - \theta) \mathbf{G}_{be}(l), \tag{10}$$

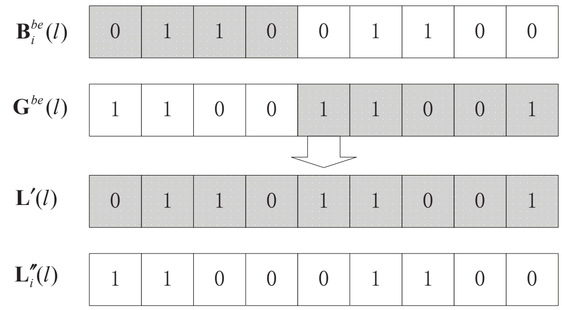


Fig. 1: $\mathbf{L}_i(l)$ producing process through single-point crossover.

where θ is a random variable between 0 and 1, $\mathbf{G}_{be}(l)$ denotes the global best position of all particles in the l-th iteration.

The particle location in BQPSO is a binary string, therefore the evolution equation is different from that of QPSO. In BQPSO, the iterative equation (8) is replaced by the procedure of inversing the value of each bit in $\mathbf{L}_i(l)$ with a probability. All bits in the same decision variable have the same inverse probability. Specifically, the bits in $\mathbf{L}_i(l)$ belong to the g-th decision variable, i.e., $\mathbf{L}_i^g(l)$, are inversed with probability $p_i^g(l)$ to obtain $\mathbf{D}_i^g(l+1)$. The $p_i^g(l)$ can be obtained as

$$b_i^g(l) = \upsilon \cdot d_H(\mathbf{M}_{be}^g(l), \mathbf{D}_i^g(l)) \cdot \ln(1/u), \tag{11}$$

$$p_i^g(l) = \begin{cases} b_i^g(l)/\lceil log_2 K \rceil & \text{if } b_i^g(l)/\lceil log_2 K \rceil < 1 \\ 1 & \text{otherwise,} \end{cases} \tag{12}$$

where, $\mathbf{M}_{be}^g(l)$ and $\mathbf{D}_i^g(l)$ are the mean best position bits and position bits belonging to the g-th decision variable, respectively. The $d_H(\cdot, \cdot)$ is a function that can obtain the Hamming distance of two input binary strings. In BQPSO, the j-th bit of the $\mathbf{M}_{be}(l)$, i.e., $M_{be}^j(l)$, is determined by the states of the j-th bit of all $\mathbf{B}_i^{be}(l)$. If more particles take on 1, the $M_{be}^j(l)$ will be 1; otherwise it is 0.

The local attractor $\mathbf{L}_i(l)$ in BQPSO can be obtained after single-point crossover or multipoint crossover process. Fig. 1 shows how to obtain the local attractor from $\mathbf{B}_i^{be}(l)$ and $\mathbf{G}_{be}(l)$ through single-point crossover process. First, randomly select a number between 1 and $N \times M \times \lceil log_2 K \rceil$, and regard it as the crossover point. Then $\mathbf{L}_i^{'}(l)$ and $\mathbf{L}_i^{''}(l)$ are obtained from the offsprings of $\mathbf{B}_i^{be}(l)$ and $\mathbf{G}^{be}(l)$. Finally, $\mathbf{L}_i^{'}(l)$ and $\mathbf{L}_i^{''}(l)$ are selected randomly as the $\mathbf{L}_i(l)$.

Based on the BQPSO, the RU allocation algorithm is developed. The detail steps are given in **Algorithm 1**.

IV. Power Allocation for Given RU Allocation

In this section, under the assumption of known RU allocation, we transform the original non-convex optimization problem to an equivalent concave optimization problem, and obtain the optimal power allocation policy by the Lagrange dual approach.

A. The Problem Transformation

Assuming that the RU allocation policy $\mathbf{A}^t$ is known, then the BS only needs to do the power allocation for different

Algorithm 1 RU Allocation for Given Power Allocation

Initialization:
a) Set population size I, the maximum iteration times $L_{iteration}^{BQPSO}$, and iteration index $l = 1$.
b) Initialize the RU allocation policy $\mathbf{A}_i(1)$, and obtain $\mathbf{D}_i(1)$ according to the relationship between $\mathbf{A}_i(1)$ and $\mathbf{D}_i(1)$.
c) Set $\mathbf{B}_i^{be}(1) = \mathbf{D}_i(1)$, and according to the fitness function choose a best position from $\mathbf{B}_i^{be}(1)$ as the $\mathbf{G}_{be}(1)$;
for $l = 1, \cdots, L_{iteration}^{BQPSO}$ **do**
  Calculates $\mathbf{M}_{be}(l)$ and $\mathbf{L}_i(l)$ according to the related rules;
  for $i = 1, \cdots, I$ **do**
    Obtain $\mathbf{D}_i(l+1)$ according to the rules aforementioned;
    Get the updated RU allocation policy $\mathbf{A}_i(l+1)$ according to $\mathbf{D}_i(l+1)$;
    Get the individual best RU allocation policy $\mathbf{A}_i^{ibe}(l)$ according to $\mathbf{B}_i^{be}(l)$;
    if $U[\mathbf{A}_i(l+1), \mathbf{P}^t] > U[\mathbf{A}_i^{ibe}(l), \mathbf{P}^t]$, **then** BS sets $\mathbf{B}_i^{be}(l+1) = \mathbf{D}_i(l+1)$; **else** sets $\mathbf{B}_i^{be}(l+1) = \mathbf{B}_i^{be}(l)$; **endif**;
    Get the individual best RU allocation policy $\mathbf{A}_i^{ibe}(l+1)$ according to $\mathbf{B}_i^{be}(l+1)$;
    Get the global best RU allocation policy $\mathbf{A}^{gbe}(l)$ according to $\mathbf{G}_{be}(l)$;
    if $U[\mathbf{A}_i^{ibe}(l+1), \mathbf{P}^t] > U[\mathbf{A}^{gbe}(l), \mathbf{P}^t]$, **then** BS sets $\mathbf{G}_{be}(l+1) = \mathbf{B}_i^{be}(l+1)$; **else** sets $\mathbf{G}_{be}(l+1) = \mathbf{G}_{be}(l)$; **endif**;
    $i = i + 1$
  end for
  $l = l + 1$
end for
Obtain the RU allocation policy $\mathbf{A}^t$ according to $\mathbf{G}_{be}(l)$.

users. Therefore, the resource allocation problem in (4) can be reduced to:

$$\begin{aligned} &\max_{\mathbf{A},\mathbf{P}} \frac{R_{tot}(\mathbf{A},\mathbf{P})}{P_{tot}(\mathbf{A},\mathbf{P})} \\ &\texttt{s.t.} \quad C3, C4, C5 \end{aligned} \tag{13}$$

Unfortunately, the optimization problem in (13) is still a non-convex optimization problem due to the C5 and the fractional form of the objective function. In order to develop an efficient resource allocation algorithm, several transformations are needed to eliminate the non-convexity and to make the problem more tractable. In the following, we first tackle the C5 by changing the independent variable, and then the original objective function is transformed to an equivalent form which is concave with respect to the new independent variable.

The C5 makes the feasible set non-convex. In general, to solve the problem efficiently one need to linearize the C5. We introduce a new independent variable $R_{n,m,k} = WT \log 2(1+\gamma_{n,m,k} p_{n,m,k})$ to the problem (13), which can decouple the proportional rate constraints. After introducing $R_{n,m,k}$, C5 can be rewritten as C5':

$$\sum_{(n,m)\in\mathbb{A}_k^t} R_{n,m,k} = \frac{\eta_k}{\eta_{K_1+1}} \sum_{(n,m)\in\mathbb{A}_{K_1+1}^t} R_{n,m,K_1+1}, \forall k \in \Omega_B. \tag{14}$$

Similarly, C3 and C4 can be rewritten as:

$$\begin{aligned} C3' &: \sum_{k=1}^{K}\sum_{n=1}^{N} \frac{2^{\frac{R_{n,m,k}}{WT}} - 1}{\gamma_{n,m,k}} \le P_{max}, \quad \forall m, \\ C4' &: \sum_{(n,m)\in\mathbb{A}_k^t} R_{n,m,k} \ge R_k^{min}, \quad \forall k \in \Omega_A, \end{aligned} \tag{15}$$

Since $R_{n,m,k}$ is a non-negative variable, it is necessary to add a new constraint C6: $R_{n,m,k} \ge 0$. Furthermore, it easy to verify that the C3' and C4' are also concave functions or affine functions with respect to $R_{n,m,k}$. Therefore, the feasible set is a convex set after the transformation.

With a convex feasible set, [28] and [29] show that the fractional program problem in (13) can be transformed to a easily solvable form. We define the maximum energy efficiency q^t of the considered system as

$$q^t = \frac{R_{tot}(\mathbf{A}^t,\mathbf{P}^t)}{P_{tot}(\mathbf{A}^t,\mathbf{P}^t)} = \max_{\mathbf{P}} \frac{R_{tot}(\mathbf{A}^t,\mathbf{P})}{P_{tot}(\mathbf{A}^t,\mathbf{P})}. \tag{16}$$

Then, we can use the following theorem which had been proved in [28] and [29].

Theorem 1. *The maximum energy efficiency q^t is achieved if and only if*

$$\begin{aligned} &\max_{\mathbf{P}} \ R_{tot}(\mathbf{A}^t,\mathbf{P}) - q^t P_{tot}(\mathbf{A}^t,\mathbf{P}) \\ &\quad = R_{tot}(\mathbf{A}^t,\mathbf{P}^t) - q^t P_{tot}(\mathbf{A}^t,\mathbf{P}^t) = 0, \end{aligned} \tag{17}$$

for $R_{tot}(\mathbf{A}^t,\mathbf{P}) \ge 0$ and $P_{tot}(\mathbf{A}^t,\mathbf{P}) > 0$.

Theorem 1 states that: for an optimization problem with a fractional form objective function, there exists an equivalent objective function in subtractive form, e.g., $R_{tot}(\mathbf{A}^t,\mathbf{P}^t) - q^t P_{tot}(\mathbf{A}^t,\mathbf{P}^t)$. When the RU allocation policy is known $p^{const} = \sum_{k=1}^{K} P_{r_k} M_{r_k}(\mathbb{A}_k) + \sum_{k=1}^{K}(M - M_{r_k}(\mathbb{A}_k))P_{nr_k}) + P_c M$ is a constant. Then the objective function can be transformed with the independent variable $R_{n,m,k}$ as:

$$\begin{aligned} U^{eff}(\mathbf{R}) = &\sum_{k=1}^{K}\sum_{(n,m)\in\mathbb{A}_k^t} R_{n,m,k} - qT\Big[P^{const} \\ &+ \sum_{k=1}^{K}\sum_{(n,m)\in\mathbb{A}_k^t} \frac{2^{\frac{R_{n,m,k}}{WT}} - 1}{\gamma_{n,m,k}}\Big] \end{aligned} \tag{18}$$

It is easily to verify that $U^{eff}(\mathbf{R})$ is a concave function with respect to $R_{n,m,k}$. As a result, the transformed problem in (19) is a concave optimization problem. Hence, we can first solve the (19), and then we can use iterative algorithms, such as the Dinkelbach method [28] to solve the (13).

$$\begin{aligned} &\max_{\mathbf{R}} U(\mathbf{R})^{eff} \\ &\texttt{s.t.} C3', C4', C5', C6. \end{aligned} \tag{19}$$

In this paper, Dinkelbach method is adopted to design the power allocation algorithm which is described in **Algorithm 2**. The proposed algorithm can converges to the optimal energy efficiency, which is proved in the Appendix. Furthermore, since $q_{l+1} = R_{tot}(\mathbf{A}^t, \mathbf{P}_l, R_{be}^l)/P_{tot}(\mathbf{A}^t, \mathbf{P}_l, R_{be}^l)$, the algorithm converges to the optimal energy efficiency with a superlinear convergence rate [9].

$$L(R_{m,n,k},\lambda_m,\beta_k,\xi_k)=\sum_{k=1}^{K}\sum_{(n,m)\in\mathbb{A}_k^t}R_{n,m,k}-qT\Big[P^{const}+\sum_{k=1}^{K}\sum_{(n,m)\in\mathbb{A}_k^t}\frac{2^{\frac{R_{n,m,k}}{WT}}-1}{\gamma_{n,m,k}}\Big]+\sum_{k=1}^{K_1}\beta_k\Big(\sum_{(n,m)\in\mathbb{A}_k^t}R_{n,m,k}$$
$$-R_k^{min}\Big)+\sum_{m=1}^{M}\lambda_m\Big(P_{max}-\sum_{k=1}^{K}\sum_{n=1}^{N}\frac{2^{\frac{R_{n,m,k}}{WT}}-1}{\gamma_{n,m,k}}\Big)+\sum_{k=K_1+2}^{K}\xi_k\Big(\sum_{(n,m)\in\mathbb{A}_k^t}R_{n,m,k}-\frac{\eta_k}{\eta_{K_1+1}}\sum_{(n,m)\in\mathbb{A}_{K_1+1}^t}R_{n,m,K_1+1}\Big),\tag{20}$$

Algorithm 2 Power Allocation for Given RU Allocation

Initialization:
a) Set the maximum iteration times $L_{iteration}^{Dinkelbach}$, and the maximum tolerance ϵ;
b) Initialize the optimal energy efficiency $q_1=0$ and iteration index $l=1$.
for $l=1,\cdots,L_{iteration}^{Dinkelbach}$ **do**
  For given q_l the BS solves the problem in (19), and obtains resource allocation policy $\{\mathbf{A}^t,\mathbf{P}_l\}$;
  if $|R_{tot}(\mathbf{A}^t,\mathbf{P}_l)-q_lP_{tot}(\mathbf{A}^t,\mathbf{P}_l)|<\epsilon$, **then** the BS obtains the power allocation policy $\mathbf{P}^t=\mathbf{P}_l$;
  else BS sets $q_{l+1}=\frac{R_{tot}(\mathbf{A}^t,\mathbf{P}_l)}{P_{tot}(\mathbf{A}^t,\mathbf{P}_l)}$ and $l=l+1$; **endif**;
end for
Output the optimal power allocation policy $\mathbf{P}^t=\mathbf{P}_l$.

B. Power Allocation for Transformed Problem

Since the optimization problem in (19) is a concave optimization problem. Thus, under some mild conditions, it can be shown that strong duality holds and the duality gap is equal to zero [30]. In other words, solving the optimization problem in (19) is equivalent to solve the Lagrange dual problem. The Lagrange function of the transformed problem is given as (20) on the top of the page, where $\lambda_m\geq 0$ ($m=1,\cdots,M$), $\beta_k\geq 0$ ($k=1,\cdots,K_1$), and $\xi_k\geq 0$ ($k=K_1+2,\cdots,K$) are the Lagrangian multipliers. When deriving the power allocation policy, the boundary constraints $p_{n,m,k}\geq 0$ and $R_{n,m,k}\geq 0$ will be absorbed into the Karush-Kuhn-Tucker (KKT) conditions. Thus, the dual problem of (19) is as follows:

$$\min_{\lambda_m,\beta_k,\xi_k}\max_{R_{m,n,k}}L(R_{m,n,k},\lambda_m,\beta_k,\xi_k).\tag{21}$$

In the following, we solve the dual problem iteratively by decomposing it into two layers: layer 1 subproblem - power allocation for fixed set of Lagrange multipliers; layer 2 master problem - obtaining the Lagrange multipliers with the gradient method.

1) Solution for Layer 1: By dual decomposition, the BS first solves the following Layer 1 subproblem

$$\max_{R_{m,n,k}}L(R_{m,n,k},\lambda_m,\beta_k,\xi_k),\tag{22}$$

with a given parameter q and a fixed set of Lagrange multipliers $\{\lambda_m,\beta_k,\xi_k\}$. Using standard optimization techniques and the KKT conditions, the power allocation policy $p_{n,m,k}$ is obtained as

$$\begin{cases}\Big[\frac{(1+\beta_k)TW}{(qT+\lambda_m)\ln 2}-\frac{1}{\gamma_{n,m,k}}\Big]^{+} & \forall k\in\Omega_A\\ \Big[\frac{(1-\sum_{K_1+2}^{K}\frac{\xi_k\eta_k}{\eta_{K_1+1}})TW}{(qT+\lambda_m)\ln 2}-\frac{1}{\gamma_{n,m,k}}\Big]^{+} & k=K_1+1\\ \Big[\frac{(1+\xi_k)TW}{(qT+\lambda_m)\ln 2}-\frac{1}{\gamma_{n,m,k}}\Big]^{+} & \forall k\in\Omega_B,\ k\neq K_1+1.\end{cases}\tag{23}$$

where $[x]^{+}=\max\{0,x\}$. The power allocation has the form of multi-level water-filling. It can be observed that the energy efficiency variable $q\geq 0$ prevents energy inefficient transmission by truncating the water-levels.

2) Solution for Layer 2: The dual function is differentiable with respect to optimization variables $R_{n,m,k}$ ($p_{n,m,k}$). Therefore, using the solutions of the Layer 1 subproblems, the gradient method [29] can be used to solve the Layer 2 master problem, which leads to

$$\lambda_m^{l+1}=\Big[\lambda_m^{l}-\nu_m^{l}\times(P_{max}-\sum_{k=1}^{K}\sum_{n=1}^{N}p_{n,m,k})\Big]^{+},\forall m,\tag{24}$$

$$\begin{aligned}\beta_k^{l+1}=&\Big[\beta_k^{l}-\vartheta_k^{l}\times(\sum_{(n,m)\in\mathbb{A}_k^t}TW\log_2(1+\\&\gamma_{n,m,k}p_{n,m,k})-R_k^{min})\Big]^{+},\forall k\in\Omega_A,\end{aligned}\tag{25}$$

$$\begin{aligned}\xi_k^{l+1}=&\Big[\xi_k^{l}+\upsilon_k^{l}\times(R_{be}\eta_k-\sum_{(n,m)\in\mathbb{A}_k^t}TW\log_2(1+\\&\gamma_{n,m,k}p_{n,m,k}))\Big]^{+},\forall k\in\Omega_B,k\neq K_1+1\end{aligned}\tag{26}$$

where index $l\geq 0$ is the iteration index, ν_m^l, ϑ_k^l and υ_k^l are positive step sizes. R_{be} is the total transmission data of all best-effort service.

$$R_{be}=\sum_{k\in\Omega_B}\sum_{(n,m)\in\mathbb{A}_k}TW\log_2(1+\gamma_{n,m,k}p_{n,m,k}),\tag{27}$$

Therefore, for each set of Lagrange multipliers $\{\lambda_m,\beta_k,\xi_k\}$, we can obtain the optimized power allocation $p_{n,m,k}$ and R_{be} from (23) and (27), respectively. After obtaining $p_{n,m,k}$ and R_{be}, we can use (24)-(26) to update the Lagrange multipliers. The process is repeated until convergence is achieved. Since the transformed problem in (19) is concave in nature, if the chosen step sizes satisfy the general conditions stated in [30], then the iteration between Layer 1 and Layer 2 will converge to the optimal solution of (19).

V. Time-Frequency-Power Resource Allocation

Based on the aforementioned works, first a joint time-frequency-power resource allocation scheme is developed, and then the complexity of the proposed scheme is analyzed.

A. Time-Frequency-Power Resource Allocation

As discussed above, firstly, when the transmission power in each RU is known, the RU allocation policy can be obtained by using the **Algorithm 1**. Secondly, based on the achieved RU allocation results, the optimal power allocation can be obtained by using the **Algorithm 2**. In the third step, we substitute the power allocation results obtained in the second step into the first step and calculate the RU allocation again. The third step is shown in **Algorithm 3**. This iteration runs repeatedly until the results converge. The initial power allocation $\mathbf{P}^1$ is uniform among all RUs.

B. Analysis of Complexity and Feasibility

In this subsection, we analyze the time complexity of the proposed time-frequency-power resource allocation scheme. First, we analyze the time complexity of the RU allocation, i.e., **Algorithm 1**. The complexity of the RU allocation is $\mathcal{O}(L_{iteration}^{BQPSO} \times I \times M \times N \times \lceil log_2 K \rceil)$ [25]. Second, we analyze the time complexity of the power allocation, i.e. **Algorithm 2**. Since the original optimization problem has been transformed to a concave problem with respect to $R_{m,n,k}$, and dual decomposition is used to obtain the power allocation policy. Therefore, similar to the analysis in [31], the complexity of the power allocation is $\mathcal{O}(L_{iteration}^{Dinkelbach} \times L_{iteration}^{Power} \times M \times N \times K)$, where $L_{iteration}^{Power}$ is the iteration time of the gradient method used to solve the Lagrange dual problem. Hence, the total complexity of the proposed resource allocation scheme is $\mathcal{O}([L_{iteration}^{JTFPR} \times M \times N \times (L_{iteration}^{BQPSO} \times I \times \lceil log_2 K \rceil + L_{iteration}^{Dinkelbach} \times L_{iteration}^{Power} \times K)])$. We find that the complexity of the proposed scheme is linear with respect to the number of time slots, users, sub-carriers and the iteration times. Therefore, if the proposed scheme has good convergence property, the time complexity of the proposed scheme is acceptable. In Section VI, the convergence performance of the proposed scheme is evaluated through the method of simulation, and we can find that it is acceptable.

VI. Performance Evaluation and Discussions

The simulation parameters are set as following. The total bandwidth, 1.08 MHz, is equally divided into $N = 72$ orthogonal sub-carriers. The scheduling period includes $M = 10$ time slots and each time slot with a duration $T = 0.5$ ms. Assume that there are $K = 5$ UEs unless otherwise noted. UE 1 and UE 2 are users with the minimum-rate requirement of 500 kbps and 750 kbps, respectively. UE 3, UE 4 and UE 5 are users with best-effort service. The channel of the kth UE is modeled as Rayleigh fading with an average CNR of $\overline{\gamma}_k$. In our simulation results, the average CNR in the horizontal axis represents the CNR of the lowest CNR UE. Unless specifically noted, $I = 3000$, $\overline{\gamma}_1 = 10\overline{\gamma}_2 = \overline{\gamma}_3 = 10\overline{\gamma}_4 = 5\overline{\gamma}_5$, $\eta_3 = \eta_4 = \eta_5 = 1/3$, $P_{max} = 40.00$ dBm, $P_c = 36.99$ dBm, $P_{r_k} = [31.14, 31.46, 30.79, 31.14, 31.46]$ dBm, and $P_{nr_k} = [20.00, 23.01, 20.00, 23.01, 23.01]$.

Algorithm 3 Time-Frequency-Power Resource Allocation

Initialization:
a) Each UE estimates the $h_{n,m,k}$, and sends $h_{n,m,k}$, energy consumption P_{r_k} and P_{nr_k} to the BS;
b) Set the maximum iteration times $T_{iteration}^{JTFPR}$, and the maximum tolerance ε;
c) The BS initializes the time-frequency RU allocation policy $\mathbf{A}^1$, the power allocation policy $\mathbf{P}^1$, and iteration index $t = 1$.
for $t = 1, \cdots, T_{iteration}^{JTFPR}$ **do**
  For given power allocation policy $\mathbf{P}^t$, the BS obtains time-frequency RU allocation policy $\mathbf{A}^{t+1}$ via **Algorithm 1**;
  After getting the RU allocation policy $\mathbf{A}^{t+1}$, the BS calculates the power allocation policy $\mathbf{P}^{t+1}$ by **Algorithm 2**;
  if $|p_{n,m,k}^{t+1} - p_{n,m,k}^{t}| \leq \varepsilon, \forall n, m, k$, **then** the BS obtains the resource allocation policy $\{\mathbf{A}^*, \mathbf{P}^*\} = \{\mathbf{A}^t, \mathbf{P}^t\}$;
  else $t = t + 1$; **endif**
end for
Output the resource allocation policy $\{\mathbf{A}^*, \mathbf{P}^*\} = \{\mathbf{A}^t, \mathbf{P}^t\}$.

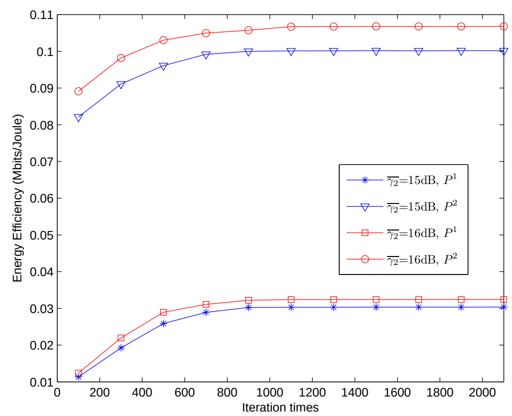


Fig. 2: Convergence of **Algorithm 1**.

A. Convergence of the Proposed Resource Allocation Scheme

Fig. 2 illustrates convergence of the proposed RU allocation algorithm for given power allocation. The given power allocation policy are $\mathbf{P}^1$ and $\mathbf{P}^2$. The results in Fig. 2 were averaged over 500 adaptation processes. It can be seen that no matter which power allocation policy and channel condition are given, the proposed RU allocation algorithm always converges to 90% of the upper bound performance within $L_{iteration}^{BQPSO} = 700$.

Fig. 3 illustrates convergence of the gradient method used to solve the Lagrange dual problem in power allocation. $\overline{\gamma}_2 = 15dB$, the RU allocation policy is $\mathbf{A}^1$ with $L_{iteration}^{BQPSO} = 1000$ and $q = 0.1Mbit/J$. $L_{iteration}^{BQPSO} = 1000$ can ensure the RU allocation algorithm convergence. It can be seen that the gradient method has fast convergence rate, it converges to 90% of the upper bound performance within 5 iterations.

Fig. 4 illustrates convergence of the proposed power allocation algorithm for given RU allocation, i.e., Dinkelbach method. The RU allocation policy is $\mathbf{A}^1$ with $L_{iteration}^{BQPSO} = 1000$, $L_{iteration}^{Power} = 10$. It can be seen that the Dinkelbach method has fast convergence rate, it converges to 90% of the upper bound performance within 6 iterations.

Fig. 5 illustrates convergence of the proposed joint time-frequency-power resource allocation scheme. $L_{iteration}^{BQPSO} =$

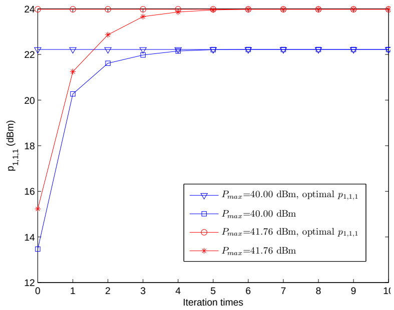


Fig. 3: Convergence of gradient method.

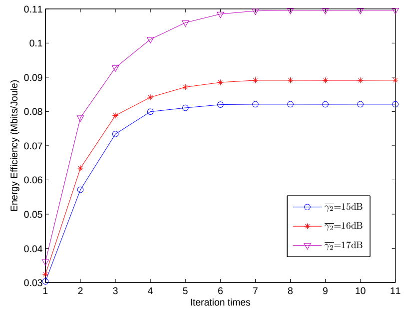


Fig. 4: Convergence of **Algorithm2**.

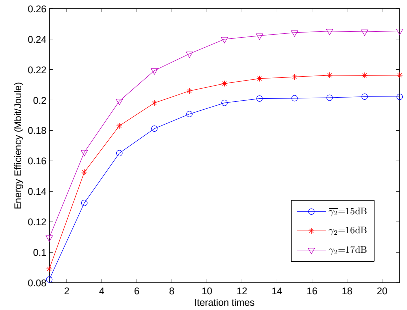
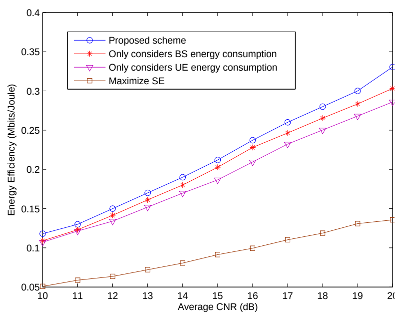
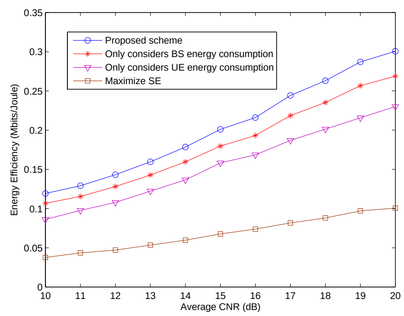


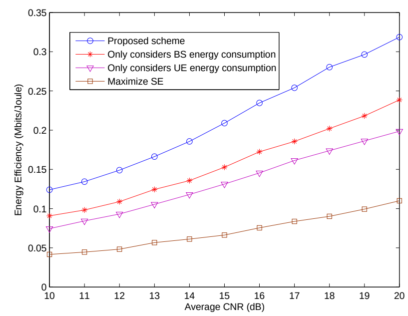
Fig. 5: Convergence of **Algorithm3**.



a) Scenario 1.



b) Scenario 2.



c) Scenario 3.

Fig. 6: Energy efficiency of different resource allocation schemes.

1000, $L_{iteration}^{Power} = 10$, $L_{iteration}^{Dinkelbach} = 10$. Similarly, it can be seen that the proposed scheme has satisfactory convergence rate, and it converges to 90% of the upper bound performance within 11 iterations.

From Fig. 2 - 5, we can find that the proposed resource allocation scheme has good convergence performance.

B. Performance Comparison of Different Resource Allocation Schemes

In order to show that the proposed scheme is necessary for some scenarios, the simulation is performed in different scenarios. The communication scenarios in practical can be roughly classified into following three scenarios. Scenario 1: The UE circuit power is little compared with BS transmission power, i.e., $P_{max} = 43.01$ dBm, $P_c = 40.00$ dBm, $P_{r_k} = [28.45, 29.03, 28.75, 29.03, 28.45]$ dBm and $P_{nr_k} = [20.00, 23.01, 20.00, 23.01, 20.00]$ dBm. Scenario 2: The UE circuit power is comparable to BS transmission power, i.e., $P_{max} = 40.00$ dBm, $P_c = 36.99$ dBm, $P_{r_k} = [31.14, 31.46, 30.79, 31.14, 31.46]$ dBm and $P_{nr_k} = [20.00, 23.01, 20.00, 23.01, 23.01]$ dBm. Scenario 3: The UE circuit power plays an important role in the total energy consumption, i.e., $P_{max} = 36.99$ dBm, $P_c = 33.01$ dBm, $P_{r_k} = [31.14, 31.76, 30.00, 30.79, 31.46]$ dBm and $P_{nr_k} = [20.00, 23.01, 20.00, 24.77, 20.00]$ dBm. Furthermore, to evaluate the proposed resource allocation scheme, we compare it with three traditional resource allocation schemes. Comparison Scheme 1: Only considers the BS energy consumption as in [9], [11]. Comparison Scheme 2: Only considers the UE energy consumption as in [7]. Comparison Scheme 3: Maximize system transmission data rate, i.e., maximize the SE.

Fig. 6 show the energy efficiency of different resource allocation schemes under the aforementioned three scenarios. Fig. 6 show that the proposed scheme can achieve the best energy efficiency performance in all the scenarios. Furthermore, the results show that when the proportion of UE circuit power to the total power become larger, i.e., from scenario 1 to scenario 3, the advantage of the proposed scheme increases. This is because all energy consumption during communication process are considered when designing the proposed scheme, which is different from the comparing schemes. Therefore, better energy efficiency performance is achieved. In the current and future communication systems, more and more communications will happen in short-range situations, and circuit energy consumption in receiver will play an important role in total energy consumption. Therefore, our work is meaningful.

Fig. 7 shows the energy efficiency of different resource allocation schemes versus the number of sub-carriers. We find that the energy efficiency rises up progressively as the number of sub-carriers increases gradually. The reason is that as the number of sub-carriers increase, more bandwidth resources are available, then better energy efficiency can be obtained, which is a classical conclusion and had obtained in [15], [16]. In addition, just like Fig. 6, we can also find that the proposed scheme can achieve the best performance of energy efficiency among all the resource allocation schemes.

Fig. 8 shows the energy efficiency of different resource allocation schemes versus the number of UEs. To simply the presentation, we assume $K_1 = \lfloor 0.4 * K \rfloor$, where $\lfloor \cdot \rfloor$ means rounding down the value. $R_k^{min} = 1.25/K_1$ Mbps, $\forall k \in \Omega_A$,

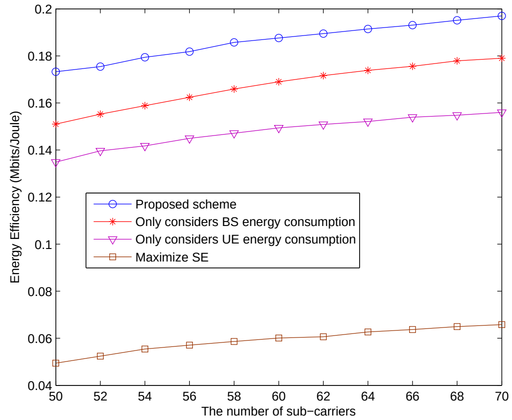


Fig. 7: Energy efficiency of different resource allocation schemes versus the number of sub-carriers ($\overline{\gamma}_2 = 15$ dB).

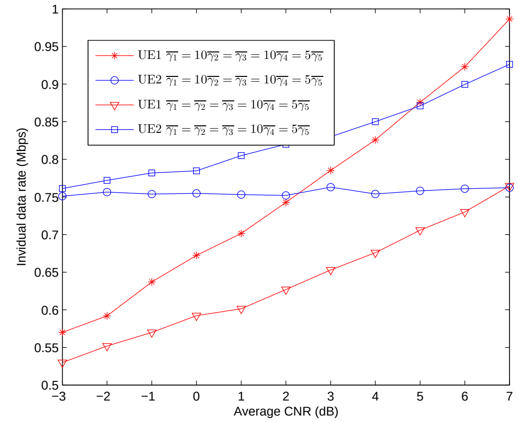


Fig. 9: Satisfying the transmission rate requirement of minimum-rate guarantee service.

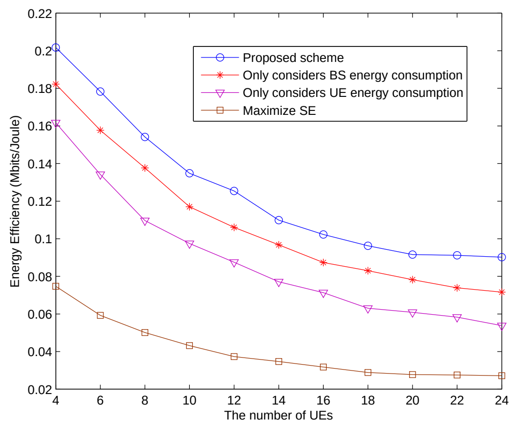


Fig. 8: Energy efficiency of different resource allocation schemes versus the number of UEs.

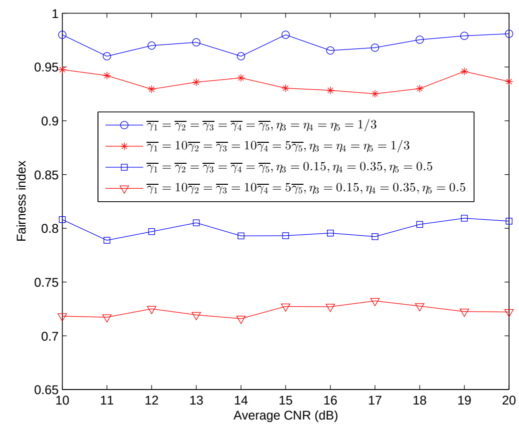


Fig. 10: Guaranteeing the fairness of best-effort service.

and $\eta_k = 1/K_2, \forall k \in \Omega_B$. The channel conditions of different UEs are independent, and all UEs have the same average CNR 15 dB. The circuit power of different UE is also set as the same, $P_{r_k} = 31.14$ dBm and $P_{nr_k} = 20.00$ dBm ($\forall k \in [1, \cdots, K]$). Furthermore, BS maximum transmission power and circuit power are set as $P_{max} = 40.00$ dBm, $P_c = 36.99$ dBm, respectively. From Fig. 8, we can obtain the conclusion that the energy efficiency decreases gradually as the number of UEs increases. Since the UE circuit power consumption is considered, and the greater the number of UEs, the more circuit power will be consumed, which results in low energy efficiency. The conclusion is different from existing results. When receiver circuit power is not considered, the nominal energy efficiency will rise gradually as the number of UEs increases, for the multi-users diversity gain. Furthermore, Fig. 8 also proves that the proposed scheme can achieve the best performance of energy efficiency.

C. Capability of the Proposed Resource Allocation Scheme for Guaranteeing Heterogeneous Service QoS

In this subsection, we discuss the performance of the proposed resource allocation scheme for guaranteeing heterogeneous QoS requirements. Fig. 9 shows the capability of the proposed scheme for satisfying the requirement of minimum-rate guarantee service under different channel conditions. From Fig. 9, we can see that the proposed scheme can guarantee users' minimum-rate requirements under all channel conditions. Furthermore, under the condition of $\overline{\gamma}_1 = 10\overline{\gamma}_2 = \overline{\gamma}_3 = 10\overline{\gamma}_4 = 5\overline{\gamma}_5$, we find that the rate of low CNR UE (UE2) is almost fixed at the minimum-rate requirement, 750 kbps, while the rate of high CNR UE (UE1) increases with the CNR. This is because, in the proposed scheme more resource is allocated to the UE with good channel condition to achieve better energy efficiency.

For the best-effort services, the fairness can be evaluated in term of fairness index [32], which is defined as

$$\phi = \frac{(\sum_{k=K_1+1}^{K} R_k)^2}{(K - K_1)\sum_{k=K_1+1}^{K} R_k^2}. \tag{28}$$

where ϕ is in the range of $[0, 1]$, and the value of ϕ more closer to 1, the better fairness performance will be achieved.

Fig. 10 shows the capability of the proposed scheme for achieving the fairness of best-effort services under different situations. We see that no matter in what channel conditions, the fairness performance of situation $\eta_3 = \eta_4 = \eta_5 = 1/3$ is always better than situation $\eta_3 = 0.15, \eta_4 = 0.35, \eta_5 = 0.5$. Furthermore, we also find that although the channel conditions of different UEs are very different in the case of $\overline{\gamma}_1 = 10\overline{\gamma}_2 = \overline{\gamma}_3 = 10\overline{\gamma}_4 = 5\overline{\gamma}_5$, if we set reasonable η_k ($\eta_3 = \eta_4 = \eta_5 = 1/3$), the proposed scheme can still obtain satisfactory fairness performance. Therefore, we have the following conclusion: the proportional-fairness factor η_k setting has a big impact on guaranteeing service fairness, and we can adjust the proportional-fairness factor to achieve the desirable performance of guaranteeing service fairness.

VII. Conclusion

In this paper, we studied the problem of EE resource allocation for downlink communications in OFDMA networks that support heterogeneous services. Both the transmitter energy consumption and receiver energy consumption are considered. We formulated the problem of EE resource allocation as a mixed combinatorial and non-convex optimization problem. To reduce the computational complexity, we solved the problem with three steps, where techniques such as BQPSO and some mathematical processes have been used. We run simulations to evaluate the performance of the proposed scheme. Our simulation results show the effectiveness of the proposed scheme. In the future work, we will consider the issues of how to design EE resource allocation scheme with much lower computational complexity while maintaining the desirable system performance; how to evaluate the performance of the proposed scheme with realistic energy consumption models.

Appendix
Proof of **Algorithm 2** Convergence

A similar approach as in [28] is adopted to prove the convergence of the iterative algorithm, i.e., **Algorithm 2**. First, two propositions are introduced, and then the convergence is demonstrated. To simplify the notational, the equivalent objective function in (13) is defined as $F_e(q^{'}) = \max_{\mathbf{P}}\{R_{tot}(\mathbf{A}^t, \mathbf{P}) - q^{'} P_{tot}(\mathbf{A}^t, \mathbf{P})\}$.

Proposition 1: $F_e(q^{'})$ is a non-negative function in the domain of definition.

Proof: Assuming $\{\mathbf{A}^t, \mathbf{P}^{'}\}$ is an arbitrary solution for the problem and $q^{'} = R_{tot}(\mathbf{A}^t, \mathbf{P}^{'})/P_{tot}(\mathbf{A}^t, \mathbf{P}^{'})$, then

$$\begin{aligned} F_e(q^{'}) &= \max_{\mathbf{P}}\{R_{tot}(\mathbf{A}^t, \mathbf{P}) - q^{'} P_{tot}(\mathbf{A}^t, \mathbf{P})\} \\ &\geq R_{tot}(\mathbf{A}^t, \mathbf{P}^{'}) - q^{'} P_{tot}(\mathbf{A}^t, \mathbf{P}^{'}) = 0. \end{aligned} \tag{29}$$

Proposition 2: $F_e(q^{'})$ is a strictly monotonic decreasing function with respect to $q^{'}$, i.e., $F_e(q^{''}) > F_e(q^{'})$ as long as $q^{'} > q^{''}$.

Proof: Assuming $\{\mathbf{A}^t, \mathbf{P}^{'}\}$ and $\{\mathbf{A}^t, \mathbf{P}^{''}\}$ are two different optimal policies for $F_e(q^{'})$ and $F_e(q^{''})$, respectively. Then

$$\begin{aligned} F_e(q^{''}) &= \max_{\mathbf{P}}\{R_{tot}(\mathbf{A}^t, \mathbf{P}) - q^{''} P_{tot}(\mathbf{A}^t, \mathbf{P})\} \\ &= R_{tot}(\mathbf{A}^t, \mathbf{P}^{''}) - q^{''} P_{tot}(\mathbf{A}^t, \mathbf{P}^{''}) \\ &> R_{tot}(\mathbf{A}^t, \mathbf{P}^{'}) - q^{''} P_{tot}(\mathbf{A}^t, \mathbf{P}^{'}) \\ &\geq R_{tot}(\mathbf{A}^t, \mathbf{P}^{'}) - q^{'} P_{tot}(\mathbf{A}^t, \mathbf{P}^{'}) = F(q^{'}). \end{aligned} \tag{30}$$

Therefore, the convergence of the **Algorithm 2** can be proved as following. First, we can prove that q increases in each iteration. Second, we demonstrate that if iteration time is large enough, q will converge to the optimal solution q^t such that it meets the optimality condition of **Theorem 1**.

Assuming $\{\mathbf{A}^t, \mathbf{P}_l\}$ is the optimal policy in the l-th iteration. And $q_l (\neq q^t)$ and $q_{l+1} (\neq q^t)$ represent the energy efficiency in iterations l and $l+1$, respectively. According to **Theorem 1** and **Proposition 1**, $F_e(q_l) > 0$ and $F_e(q_{l+1}) > 0$ must be true. Moreover, Since we calculate q_{l+1} as $q_{l+1} = R_{tot}(\mathbf{A}^t, \mathbf{P}_l)/P_{tot}(\mathbf{A}^t, \mathbf{P}_l)$, the $F_e(q_l)$ can be expressed as

$$\begin{aligned} F_e(q_l) &= R_{tot}(\mathbf{A}^t, \mathbf{P}_l) - q_l R_{tot}(\mathbf{A}^t, \mathbf{P}_l) \\ &= R_{tot}(\mathbf{A}^t, \mathbf{P}_l)(q_{l+1} - q_l) > 0. \end{aligned} \tag{31}$$

Since $R_{tot}(\mathbf{A}^t, \mathbf{P}_l) > 0$, then $q_{l+1} > q_l$.

Therefore, according to **Proposition 1**, **Proposition 2**, and $q_{l+1} > q_l$, as long as the iteration time is large enough, $F_e(q_l)$ will eventually approach to zero and satisfy the optimality condition of **Theorem 1**, i.e., $F_e(q^t) = 0$.

References

[1] D. Feng, C. Jiang, G. Lim, J. Cimini, L. J., G. Feng, and G. Y. Li, "A survey of energy-efficient wireless communications," *IEEE commun. surveys & tutorials*, vol. 15, no. 1, pp. 167–178, Feb. 2013.

[2] Z. Hasan, H. Boostanimehr, and V. K. Bhargava, "Green cellular networks: a survey, some research issues and challenges," *IEEE commun. surveys & tutorials*, vol. 13, no. 4, pp. 524–540, Nov. 2011.

[3] G. Y. Li, Z. Xu, C. Xiong, C. Yang, S. Zhang, Y. Chen, and S. Xu, "Energy-efficient wireless communication tutorial, survey, and open issues," *IEEE Wireless Commun.*, vol. 18, no. 6, pp. 28–35, Dec. 2011.

[4] Y. Chen, S. Zhang, S. Xu, and G. Li, "Fundamental trade-offs on green wireless networks," *IEEE Commun. Mag.*, vol. 49, no. 6, pp. 30–37, Jun. 2011.

[5] S. Chen, Y. Wang, W. Ma, and J. Chen, "Technical innovations promoting standard evolution: from TD-SCDMA to TD-LTE and beyond," *IEEE Wireless Communications*, vol. 19, no. 1, pp. 60–66, Feb. 2012.

[6] S.Pollin, R. Mangharam, B. Bougard, L. V. der Perre, R. R. I.Moerman, and F. Catthoor, "MEERA: cross-layer methodology for energy efficient resource allocation in wireless networks," *IEEE Trans. Wireless Commun.*, vol. 7, no. 1, pp. 98–109, Jan. 2008.

[7] F.-S. Chu, K.-C. Chen, and G. Fettweis, "Green resource allocation to minimize receiving energy in OFDMA cellular systems," *IEEE commun. lett.*, vol. 16, no. 3, pp. 372–374, Mar. 2012.

[8] C. Xiong, G. Y. Li, S. Zhang, Y. Chen, and S. Xu, "Energy-efficient resource allocation in OFDMA networks," in *Proc. IEEE Global Telecommunications Conference (GLOBECOM)*, Dec. 2011, pp. 1–5.

[9] C. Isheden, Z. Chong, E. A. Jorswieck, and G. Fettweis, "Framework for link-level energy efficiency optimization with informed transmitter," *IEEE Trans. Wireless Commun.*, vol. 11, no. 8, pp. 2946–2957, Aug. 2012.

[10] C. Isheden and G. P. Fettweis, "Energy-efficient multi-carrier link adaptation with sum rate-dependent circuit power," in *Proc. IEEE Global Telecommunications Conference (GLOBECOM)*, Dec. 2010, pp. 1–6.

[11] G. Miao, N. Himayat, and G. Y. Li, "Energy-efficient link adaptation in frequency-selective channels," *IEEE Trans. Commun.*, vol. 58, no. 2, pp. 545–554, Feb. 2010.

[12] C. Xiong, G. Y. Li, S. Zhang, Y. Chen, and S. Xu, "Energy- and spectral-efficiency tradeoff in downlink OFDMA networks," *IEEE Trans. Wireless Commun.*, vol. 10, no. 1, pp. 3874–3886, Nov. 2011.

[13] W. Cheng, X. Zhang, and H. Zhang, "Joint spectrum and power efficiencies optimization for statistical QoS provisionings over SISO/MIMO wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 5, pp. 903–915, May 2013.

[14] L. Zhou, R. Q. Hu, Y. Qian, and H.-H. Chen, "Energy-spectrum efficiency tradeoff for video streaming over mobile ad hoc networks," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 5, pp. 981–991, May 2013.

[15] B. Changhun and W. E. Stark, "End-to-end energy-bandwidth tradeoff in multihop wireless networks," *IEEE Trans. Inf. Theory*, vol. 55, no. 9, pp. 4051–4066, Sept. 2009.

[16] C. L. Chen, W. E. Stark, and S. G. Chen, "Energy-bandwidth efficiency tradeoff in MIMO multi-hop wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 29, no. 8, pp. 1537–1546, Sept. 2011.

[17] C. Xiong, G. Y. Li, Y. Liu, and S. Xu, "Qos driven energy-efficient design for downlink OFDMA networks," in *Proc. IEEE Global Communications Conference (GLOBECOM)*, Dec. 2012, pp. 4320–4325.

[18] F. Meshkati, H. V. Poor, S. C. Schwartz, and R. V. Balan, "Energy-efficient resource allocation in wireless networks with quality-of-service constraints," *IEEE Trans. Commun.*, vol. 57, no. 11, pp. 3406–3414, Nov. 2009.

[19] R. Gupta and E. C. Strinati, "Green scheduling to minimize base station transmit power and UE circuit power consumption," in *Proc. IEEE International Symposium on Personal Indoor and Mobile Radio Communications (PIMRC)*, Sept. 2011, pp. 2424–2429.

[20] S. G. Cui, A. J. Goldsmith, and A. Bahai, "Energy-constrained modulation optimization," *IEEE Trans. on Wireless Commun.*, vol. 4, no. 5, pp. 2349–2360, Sept. 2005.

[21] I. Stanojev, O. Simeone, Y. Bar-Ness, and H. K. Dong, "Energy efficiency of non-collaborative and collaborative hybrid-ARQ protocols," *IEEE Trans. Wireless Commun.*, vol. 8, no. 1, pp. 326–335, Jan. 2009.

[22] B. Huang, H. Tian, L. Chen, and J. Zhu, "DRX-aware scheduling method for delay-sensitive traffic," *IEEE Commun. Lett.*, vol. 14, no. 12, pp. 1113–1115, Dec. 2010.

[23] G. Miao, N. Himayatyy, and G. Y. Li, "Cross-layer optimization for energy-efficient wireless communications: a survey," *Wireless Commun. Mobile Comput.*, vol. 9, no. 4, pp. 529–542, Apr. 2009.

[24] Q. Xu, X. Li, and H. Ji, "Multiple resource allocation in OFDMA downlink networks: End-to-end energy-efficient approach," in *Accepted by Proc. IEEE International Conference on Communications (ICC)*, Jun. 2014.

[25] M. L. Xi, J. Sun, and Y. Wu, "Quantum-behaved particle swarm optimization with binary encoding," *Control and Decision*, vol. 25, no. 1, pp. 99–104, Jan. 2010.

[26] R. Xie, F. R. Yu, and H. Ji, "Dynamic resource allocation for heterogeneous services in cognitive radio networks with imperfect channel sensing," *IEEE Trans. Veh. Technol*, vol. 61, no. 2, pp. 770–780, Feb. 2012.

[27] Y. S. Zhao, X. Li, Y. Li, and H. Ji, "Resource allocation for high-speed railway downlink MIMO-OFDM system using quantum-behaved particle swarm optimization," in *Proc. IEEE International Conference on Communications (ICC)*, Jun. 2013, pp. 936–940.

[28] W. Dinkelbach, "On nonlinear fractional programming," *Management Science*, vol. 13, pp. 492–498, Mar. 1967.

[29] D. W. K. Ng, E. Lo, and R. Schober, "Energy-efficient resource allocation in OFDMA systems with large numbers of base station antennas," *IEEE Trans. on Wireless Commun.*, vol. 11, no. 9, pp. 3292–3304, Sept. 2012.

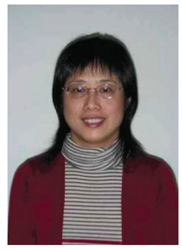
[30] S. Boyd and L. Vandenberghe, *Convex Optimization*. C ambridge, U.K.: Cambridge Univ. Press, 2004.

[31] D. W. K. Ng, E. S. Lo, and R. Schober, "Dynamic resource allocation in MIMO-OFDMA systems with full-duplex and hybrid relaying," *IEEE Trans. Commun.*, vol. 60, no. 5, pp. 1291–1304, May 2012.

[32] E. L. Hahne, "Round-robin scheduling for max-min fairness in data networks," *IEEE J. Sel. Areas Commun.*, vol. 9, no. 7, pp. 1024–1039, Sept. 1991.

Quansheng Xu received the B.S. degree with the College of Communication Engineering, Hangzhou Dianzi University, China, in 2010. He is currently pursuing his Ph.D. degree with the School of Information and Communication Engineering, Beijing University of Posts and Telecommunications, China. He served as a Reviewer of several journals and conference proceedings, including Wiley Wireless Communications and Mobile Computing, Wiley International Journal of Communication Systems and Networking, China communications, and IEEE ICC 2014. His research interests mainly include green communication, high-speed railway communication, heterogeneous networks.

Xi Li is an associate professor in the School of Information and Communication Engineering, Beijing University of Posts and Telecommunications. She received her B.E. degree and Ph.D. degree from Beijing University of Posts and Telecommunications in the major of communication and information system in 2005 and 2010 respectively. She has served as the chair of special track on cognitive testbed in Chinacom 2011, TPC member of track on MAC layer in IEEE WCNC 2012/2013 and track on MAC and cross layer design in PIMRC 2012. Her current research interests include broadband wireless communications and heterogeneous networks.



Hong Ji (SM'09) received his B.S. degree in communications engineering and the M.S. and Ph. D degrees in information and communications engineering from the Beijing university of Posts and Telecommunications (BUPT), Beijing, China, in 1989, 1992, and 2002, respectively. From June to December 2006, she was a Visiting Scholar with the University of British Columbia, Vancouver, BC, Canada. She is currently a Professor with BUPT. She is also engaged with national science research projects, including Hi-Tech Research and Development Program of China (863 program) and the National Natural Science Foundation of China. Her research interests include heterogeneous networks, peer-to-peer protocols, cognitive radio networks, relay networks, LTE/5G, and cooperative communication.

Dr. Ji serves on the editorial boards of several journals, including the Wiley International Journal of Wireless Communications and Mobile Computing, the Wiley International Journal of Communication Systems. She has also served on the Technical Program Committee (TPC) of numerous conferences, e.g., as a TPC of the 2010/2011/ 2012/2013 IEEE International Conference on Communications, the 2010/2011/2012/2013 IEEE Global Communications Conference, the 2012/2014 IEEE WCNC, the 2010 /2011 IEEE INFOCOM workshop, the 2012 IEEE CITS, and as a TPC Cochairs of the 2011 CHINACOM Workshops, a program co-chair of WINSYS 2012/2013/2014.

Dr. Xiaojiang (James) Du (SM'03) is currently an associate professor in the Department of Computer and Information Sciences at Temple University. Dr. Du received his B.S. and M.S. degree in electrical engineering from Tsinghua University, Beijing, China in 1996 and 1998, respectively. He received his M.S. and Ph.D. degree in electrical engineering from the University of Maryland College Park in 2002 and 2003, respectively. Dr. Du was an Assistant Professor in the Department of Computer Science at North Dakota State University between August 2004 and July 2009, where he received the Excellence in Research Award in May 2009. His research interests are security, cloud computing, wireless networks, computer networks and systems. He has published over 120 journal and conference papers in these areas. Dr. Du has been awarded more than \$3M research grants from the US National Science Foundation (NSF), Army Research Office, Air Force Research Lab, NASA, the Commonwealth of Pennsylvania, and Amazon. He serves on the editorial boards of four international journals. Dr. Du will serve as the Lead Chair of the Communication and Information Security Symposium of the IEEE ICC 2015, and a Co-Chair of the Mobile and Wireless Networks Track of the IEEE WCNC 2015. He was the Chair of the Computer and Network Security Symposium of the IEEE/ACM International Wireless Communication and Mobile Computing conference 2006-2010. He is (was) a Technical Program Committee (TPC) member of several premier ACM/IEEE conferences such as INFOCOM (2007-2015), IM, NOMS, ICC, GLOBECOM, WCNC, BroadNet, and IPCCC. Dr. Du is a Senior Member of IEEE and a Life Member of ACM.