Energy-efficient uplink resource allocation for IEEE 802.16j transparent-relay networks

Jia-Ming Liang [a,*], You-Chiun Wang [b], Jen-Jee Chen [c], Jui-Hsiang Liu [a], Yu-Chee Tseng [a]

[a] Department of Computer Science, National Chiao-Tung University, Hsin-Chu 30010, Taiwan
[b] Department of Computer Science and Engineering, National Sun Yat-Sen University, Kaohsiung 80424, Taiwan
[c] Department of Electrical Engineering, National University of Tainan, Tainan 70005, Taiwan

ABSTRACT

The IEEE 802.16j standard is defined to enhance WiMAX networks with relay capacity. Under the *transparent mode*, existing studies only target improving network throughput by increasing the transmission rates of *mobile stations* (*MSs*) and maximizing concurrent transmissions. However, using a higher transmission rate or allowing more concurrent transmissions could harm MSs in terms of their energy consumption, especially when they are battery-powered. In this paper, we consider the energy-conserved resource allocation problem in the uplink direction of an IEEE 802.16j network under the transparent mode. This problem asks how to arrange the frame usage with satisfying MSs' demands as the constraint and minimizing their total energy consumption as the objective. We prove this problem to be NP-complete and develop two energy-efficient heuristics, called *demand-first allocation* (*DFA*) and *energy-first allocation* (*EFA*) schemes. These heuristics employ a gradient-like search method to approximate the optimal solution. Specifically, DFA tries to satisfy MSs' demands first by using as less frame space as possible. Then, with the remaining frame space, DFA tries to save MSs' energy by lowering their transmission rates or adjusting their transmission paths. Viewed from a different perspective, EFA first allocates the frame space to MSs to consume the least energy. Since the total allocation may exceed the frame space, EFA then exploits spatial reuse and rate adjustment to pack all demands into one frame. Simulation results show that our heuristics can approximate the ideal performance bounds and save up to 90% of MSs' energy as compared to existing results.

1. Introduction

The IEEE 802.16 standard is proposed to support broadband wireless access in the emerging 4G systems. The physical layer employs the *OFDMA* (*orthogonal* frequency division multiple access) technique, where a *base station* (*BS*) can communicate with multiple *mobile stations* (*MSs*) simultaneously through a set of orthogonal channels. A typical operation of 802.16 is the *PMP* (*point-to-multi-point*) mode [1]. Recently, to overcome the coverage hole, shadow, and NLOS (non-line-of-sight) limitations, the 802.16j extension [2] is proposed to add *relay stations* (*RSs*). It has been proved in [3–5] that MSs can enjoy higher throughput and/or lower energy consumption with the help of RSs. The standard defines two types of RSs. An RS is called *transparent* if MSs are not aware of its existence. Otherwise, it is *non-transparent*. Transparent RSs are considered easier to implement than non-transparent ones since they do not need to manage the resources of networks [6].

In this paper, we consider the uplink communication in an IEEE 802.16j network with only transparent RSs (which is called a *transparent-relay network*). Given the traffic

* Corresponding author.
  E-mail addresses: jmliang@cs.nctu.edu.tw (J.-M. Liang), ycwang@cse.nsysu.edu.tw (Y.-C. Wang), jjchen@mail.nutn.edu.tw (J.-J. Chen), juihsiang@cs.nctu.edu.tw (J.-H. Liu), yctseng@cs.nctu.edu.tw (Y.-C. Tseng).

demand of each MS per frame, we consider an energy and resource allocation problem satisfying MSs' demands as the constraint and minimizing their total energy consumption as the objective. Minimizing energy consumption of MSs is critical since they are usually battery-powered. By adaptively adjusting the transmission rates of MSs and exploiting the RSs to relay data, we could reduce MSs' energy consumption. Note that when the network is working under a non-saturated condition, more unused frame space may remain. In this case, we can exploit the unused space to help further reduce the energy consumption. This will be discussed in Section 3.1.

In the literature, several studies [7–10] evaluate the network capacity of an IEEE 802.16j network. Refs. [11–14] address the placement of RSs to improve the network performance. Refs. [15–17] discuss the selection of RSs to enhance the network capacity. For transparent-relay networks, Sundaresan and Rangarajan [18] show how to leverage channel diversity and concurrent transmissions to increase network throughput. Ref. [19] suggests reusing frequency and placing RSs in an irregular manner to improve network throughput. In [20], a Markov decision process is used for admission control and a chance-constrained assignment scheme is proposed to minimize the number of RSs required and to maximize their rates. An isolation band around each RS cluster is adopted in [21] to allow more frequency reuse between RSs and the BS. Refs. [22,23] adopt a minimal coloring approach to maximize downlink capacity while reducing the differences among MSs' rates. The above studies all aim at improving network capacity but do not consider the energy conservation of MSs. A solution of a multiple-choice knapsack problem is exploited in [24] to reduce the energy consumption of MSs, but it considers the PMP mode and does not exploit RSs to help save MSs' energy.

As can be seen, existing works have not addressed the energy conservation issue well in IEEE 802.16j networks. We try to minimize MSs' energy consumption subject to satisfying their traffic demands in each frame by selecting proper paths, rates (in terms of *modulation and coding schemes*, or MCSs in short), and spatial reuse. We show this problem to be NP-complete and propose two energy-efficient heuristics, called *demand-first allocation* (*DFA*) and *energy-first allocation* (*EFA*) schemes. These two schemes try to find suboptimal solutions by exploiting a gradient-like search method. The rationale of DFA is to first find a feasible solution which uses the minimal frame space as the starting point. This implies that MSs will transmit at their maximum power levels. Then, DFA tries to lower their total energy consumption by exploiting the free frame space. On the other hand, EFA first relaxes the frame space constraint to start from a low energy solution where each MS transmits at a lower rate with no concurrent transmission. However, this may not meet all MSs' demands. Therefore, EFA tries to increase their rate/power to pack all demands into one frame. Both DFA and EFA have an iterative process to gradually improve their solutions to approximate the optimal one.

Major contributions of this paper are threefold. First, this is the first work addressing the energy and resource optimization issue in an IEEE 802.16j transparent-relay

network. In addition, we reveal that there are three key factors jointly affecting the performance of energy conservation. They are MSs' MCSs, uplink paths, and concurrent transmissions. We also do experiments and discussions to show energy consumption and resource usage are exchangeable by adjusting the three factors. Second, we prove such an energy conservation problem to be NP-complete by reducing it to the *multiple choice knapsack* problem. Then, we design two heuristics, DFA and EFA, to allocate resource for each MS. The idea is similar to the *gradient search process* [25]. DFA and EFA first find the least space cost and least energy consumption solutions, respectively, and then exploit the gradient-like search method to make the initial solutions quickly approach the optimal salutation. Third, we define two ideal bounds, demand satisfaction ratio upper bound and energy consumption lower bound, to evaluate the efficiency of the proposed schemes. Extensive simulations show that our heuristics can approximate the ideal bounds and save up to 90% of MSs' energy as compared to existing results.

The rest of this paper is organized as follows. Preliminaries are given in Section 2. Section 3 presents our energy-efficient heuristics. Simulation results are given in Section 4. Conclusions are drawn in Section 5.

2. Preliminaries

2.1. Network model

In an 802.16j transparent-relay network, there is one BS supporting multiple MSs, as shown in Fig. 1. The coverage range of the BS is defined as the reachable area when the lowest MCS (such as QPSK1/2) and the largest power are used. Inside the coverage range, RSs are deployed to help relay data between MSs and the BS. An MS can send its data to the BS either directly or indirectly through an RS. However, there are no communication links between two RSs and two MSs. Therefore, the network is a two-level tree with the BS as the root and MSs as the leaves. The standard defines two types of links for uplink communication. A link is called an *access link* if it connects to an MS at one end; otherwise, it is called a *relay link*. Fig. 1 shows some examples.

The network resource is divided into *frames*, where a frame is a two-dimensional (subchannel × time slot) array. Each frame is further divided into a *downlink subframe* and an *uplink subframe*. We show the uplink subframe in Fig. 2. It is divided into an *access zone* and a *relay zone*, which are designed for access links and relay links, respectively. The access zone is further divided into an MS–BS region and an MS–RS region. For convenience, the relay zone is also called the RS–BS region. Note that these regions have no overlap with each other. However, their sizes can be changed frame by frame.

In this work, we adopt the *PUSC* (*partial usage of subcarrier*) *mode*, which is very suitable for mobile applications [26]. Under the PUSC mode, *bursts* are the basic resource allocation units, where a burst is a sequence of slots arranged in a row-wise manner, as shown in Fig. 2. Note that a burst may cross multiple subchannels. The transmission power and rates of MSs and RSs are adjustable. However,
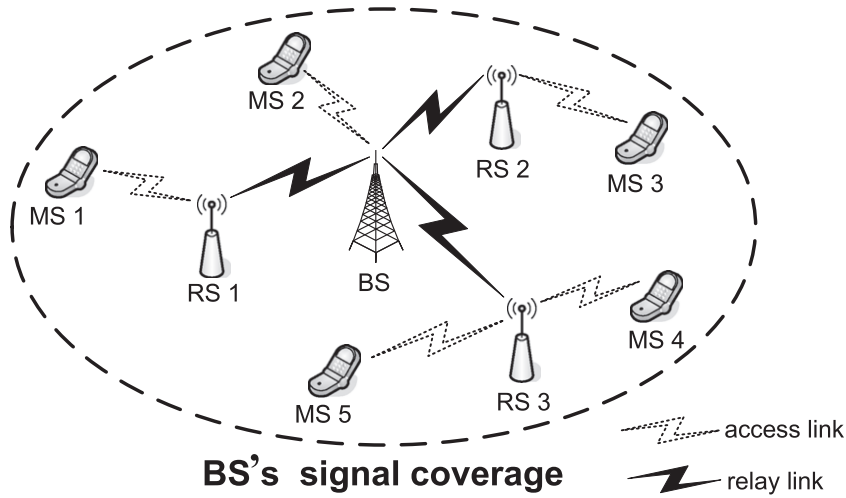
Fig. 1. The uplink communication of an 802.16j transparent-relay network.
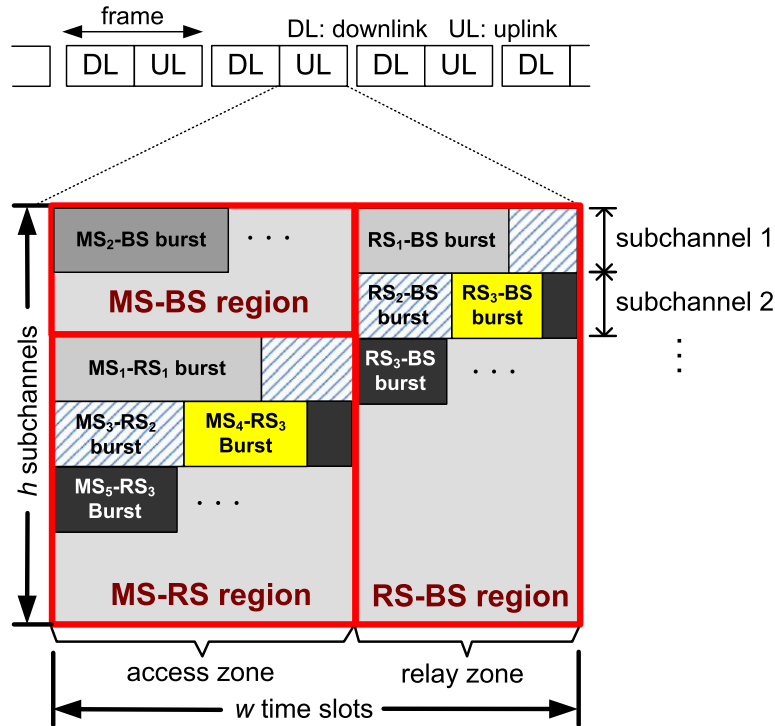

Fig. 2. The structure of the uplink subframe.

the transmission rate of an MS within one burst should be fixed. The BS is responsible for allocating bursts for MSs and RSs. In MS–BS and RS–BS regions, since the BS is the only receiver, no two bursts can overlap. In the MS–RS region, however, spatial reuse is allowed.

2.2. Energy cost model

Table 1 shows the available MCSs in IEEE 802.16j and their rates and required SINRs, denoted by $rate(\cdot)$ and $\delta(\cdot)$,

Table 1
MCSs supported by IEEE 802.16j.

MCS	scheme	$rate(MCS_k)$	$\delta(MCS_k)$
MCS_1	QPSK 1/2	48 bits/slot	6 dBm
MCS_2	QPSK 3/4	72 bits/slot	8.5 dBm
MCS_3	16QAM 1/2	96 bits/slot	11.5 dBm
MCS_4	16QAM 3/4	144 bits/slot	15 dBm
MCS_5	64QAM 2/3	192 bits/slot	19 dBm
MCS_6	64QAM 3/4	216 bits/slot	21 dBm

respectively. Let d_i be the number of bits to be transmitted by MS_i in a frame. If MS_i adopts MCS_k, then it requires $T_i = \left\lceil \frac{d_i}{rate(MCS_k)} \right\rceil$ slots to transmit its data. So, the energy cost of MS_i is $E_i = T_i \times P_i$, where P_i is the required transmission power (in mW). Suppose that there are n MSs to be served. The total energy cost is

$$E_{\text{total}} = \sum_{i=1}^{n} E_i.$$

The transmission power P_i is modeled as follows. Consider any receiver j (which can be any RS or the BS). With power P_i, the received signal power at receiver j is

$$\widetilde{P}(i,j) = \frac{G_i \cdot G_j \cdot P_i}{L(i,j)}, \tag{1}$$

where G_i and G_j are the antenna gains at MS_i and receiver j, respectively, and $L(i,j)$ is the path loss from MS_i to receiver j. Here, we adopt the standard *two-ray ground model* [27] to calculate $L(i,j)$, which is recommended by the 802.16j task group. So, the SINR (in dBm) perceived by receiver j is

$$SINR(i,j) = 10 \cdot \log_{10}\left(\frac{\widetilde{P}(i,j)}{B \cdot N_o + I(i,j)} \right), \tag{2}$$

where B is the effective channel bandwidth (in Hz), N_o is the thermal noise level, and $I(i,j)$ is the interference caused by other transmitters, which is evaluated by

$$I(i,j) = \sum_{l \neq i} \widetilde{P}(l,j).$$

MS_i's data can be correctly decoded by receiver j if

$$SINR(i,j) \geqslant \delta(MCS_k). \tag{3}$$

By integrating Eqs. (1) and (2) into Eq. (3), the minimum power required for MS_i to reach receiver j using MCS_k is

$$P_i \geqslant \frac{10^{\frac{\delta(MCS_k)}{10}}(B \cdot N_o + I(i,j)) \cdot L(i,j)}{G_i \cdot G_j}. \tag{4}$$

2.3. Problem definition

We are given an 802.16j network containing one BS, m RSs, and n MSs. Each MS_i, $i = 1, \ldots, n$, has a maximum transmission power of P_i^{MAX} (mW per subchannel) and has an uplink traffic demand of d_i bits per frame granted by the traffic management of the BS.[1] We assume that MSs may move around within the BS's signal coverage, but the relative distances among BS, RSs, and MSs can be estimated,[2] from which we can construct the network topology $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ is the node set and $\mathcal{E}$ is the communication link set. A *path* on $\mathcal{G}$ can be either a direct link from an MS to the BS or a link from an MS to an RS and then to the BS. An uplink frame has h subchannels and w time slots. Bursts in the MS–RS region can overlap with each other so as to exploit

spatial reuse. But, bursts in the MS–BS and RS–BS regions cannot overlap. If there is a burst allocated in the MS–RS region, a "matching" burst must be allocated in the RS–BS region to relay the former data. For example, in Fig. 2, since an MS_1-RS_1 burst is allocated in the MS–RS region, there must be a corresponding RS_1-BS burst allocated in the RS–BS region. However, the sizes of these two bursts may not be the same because they may use different MCSs.

Let $\mathcal{R}$ be the set of all possible paths on $\mathcal{G}$. The *energy-conserved resource allocation (ERA)* problem asks how to find a set of transmission paths $\mathcal{R}_p \subseteq \mathcal{R}$ and the corresponding MCSs, bursts, and transmission power for MSs under an $h \times w$ frame space constraint such that the total energy cost E_{total} is minimized. Specifically, we denote by $s_i = (RS_{J(i)}, MCS_{K(i)}, P_i)$ the *transmission schedule* of MS_i in a frame, where $J(i) = 0, \ldots, m$ and $K(i) = 1, \ldots, 6$. For ease of presentation, we use RS_0 as a special case to represent the BS. So, when $J(i) = 0$, it means that MS_i transmits to the BS directly using $MCS_{K(i)}$ with power P_i; otherwise, it means that MS_i transmits to $RS_{J(i)}$ using $MCS_{K(i)}$ with power P_i and then $RS_{J(i)}$ relays the data to the BS using the best possible MCS. In either case, P_i has to be bounded between the minimum required power and P_i^{MAX}, i.e.,

$$\frac{10^{\frac{\delta(MCS_{K(i)})}{10}}(B \cdot N_o + I(i, J(i))) \cdot L(i, J(i))}{G_i \cdot G_{J(i)}} \leqslant P_i \leqslant P_i^{MAX}. \tag{5}$$

In addition, we use $\mathcal{T} = \{\tau_1, \tau_2, \ldots, \tau_G\}$ to denote the set of transmission groups in a frame. Each $\tau_\ell \in \mathcal{T}$ is a transmission group consisting of either one MS–BS transmission schedule or multiple MS–RS transmission schedules. When there are multiple schedules in τ_ℓ, it means that the MSs therein can concurrently transmit to their corresponding RSs with overlapping (however, the corresponding RS–BS transmissions cannot overlap with each other). Let $\mathcal{B}_\ell$ be the binary indicator such that $\mathcal{B}_\ell = 1$ if τ_ℓ contains a single MS–BS transmission and $\mathcal{B}_\ell = 0$ otherwise. Assume that s_a is a transmission schedule in group τ_ℓ, i.e., $s_a \in \tau_\ell$. Then, the total number of slots required by the transmission group τ_ℓ is expressed by

$$S_{tot}(\tau_\ell) = \begin{cases} \left\lceil \frac{d_a}{rate(MCS_{K(a)})} \right\rceil, & \text{if } \mathcal{B}_\ell = 1, \\ S_g(\tau_\ell) + \sum_{\forall s_a \in \tau_\ell} \left\lceil \frac{d_a}{rate(MCS_{\widehat{K}(a)})} \right\rceil, & \text{if } \mathcal{B}_\ell = 0. \end{cases}$$

In the case of $\mathcal{B}_\ell = 1$, it is the number of required slots in the MS–BS region. In the case of $\mathcal{B}_\ell = 0$, it is the required slots in the MS–RS region plus those in the RS–BS region. Here, $MCS_{\widehat{K}(a)}$ is the best feasible MCS level for $RS_{J(a)}$ to relay MS_a's data to the BS. $S_g(\tau_\ell)$ is the required slots of τ_ℓ in the MS–RS region. Because of concurrent transmissions, we can conduct $S_g(\tau_\ell)$ as follows:

$$S_g(\tau_\ell) = \max_{\forall s_a \in \tau_\ell} \left\{ \left\lceil \frac{d_a}{rate(MCS_{K(a)})} \right\rceil \right\}.$$

Because the total required slots of all transmission schedules cannot exceed the frame space, we have

$$\sum_{\tau_\ell \in \mathcal{T}} S_{tot}(\tau_\ell) \leqslant h \times w. \tag{6}$$

---

[1] Here, we assume that the BS has a traffic scheduler and admission controller to manage MSs' demands according to their QoS requirements.

[2] The relative distances between MS and RSs/BS can be estimated periodically by RSs and the BS through some existing techniques such as evaluating the *received signal strength (RSS)* [28,29].

The goal of the ERA problem is to minimize the total energy consumption of all MSs:

$$\min_{s_i, i=1,\ldots,n} E_{total} = \sum_{i=1,\ldots,n} T_i \cdot P_i = \sum_{i=1,\ldots,n} \left\lceil \frac{d_i}{rate(MCS_{K(i)})} \right\rceil \cdot P_i, \quad (7)$$

by calculating the transmission schedule s_i for each MS$_i$ and group τ_ℓ that s_i belongs to, under the power constraint in Eq. (5) and the frame space constraint in Eq. (6).

Theorem 1. *The ERA problem is NP-complete.*

Proof. To simplify the proof, we consider the case of no spatial reuse in the MS–RS region and each MS has only one fixed transmission power. So, the MCS and burst(s) of each path are unique. Thus, the energy cost of an MS on each path is uniquely determined. Then, we formulate the resource allocation problem as a decision problem: *Energy-conserved resource allocation decision (ERAD)* problem: Given the network topology $\mathcal{G}$ and the demand of each MS, we ask whether or not there exists a path set $\mathcal{R}_p$ on $\mathcal{G}$ such that all MSs can conserve the total amount of energy $\mathcal{Q}$ to satisfy their demands. Then, we show ERAD problem is NP-complete.

We first show that the ERAD problem belongs to NP. Given a problem instance and a solution containing the path set, it can be verified whether or not the solution is valid in polynomial time. Thus, this part is proved.

We then reduce the *multiple-choice knapsack (MCK) problem* [30], which is known to be NP-complete, to the ERAD problem. Consider that there are n disjointed classes of objects, where each class i contains N_i objects. In each class i, every object $x_{i,j}$ has a profit $q_{i,j}$ and a weight $u_{i,j}$. Besides, there is a knapsack with capacity of $\mathcal{U}$. The MCK problem asks whether or not we can select exact one object from each class such that the total object weight is no larger than $\mathcal{U}$ and the total object profit is $\mathcal{Q}$.

We then construct an instance of the ERAD problem as follows. Let n be the number of MSs. Each MS$_i$ has N_i paths to the BS. When MS$_i$ selects a path $x_{i,j}$, it will conserve energy of $q_{i,j}$[3] and the system should allocate burst(s) of a total size of $u_{i,j}$ to transmit MS$_i$'s data to the BS. The total frame space is $w \cdot h = \mathcal{U}$. Our goal is to let all MSs conserve energy of $\mathcal{Q}$ and satisfy their demands. We show that the MCK problem has a solution if and only if the ERAD problem has a solution.

Suppose that we have a solution to the ERAD problem, which is a path set $\mathcal{R}_p$ with MSs' conserved energy and burst allocations. Each MS can choose exact one path and we need to assign paths to all MSs to satisfy their demands. The total size of bursts cannot exceed the frame space $\mathcal{U}$ and the conserved energy of all MSs is $\mathcal{Q}$. By viewing the paths of an MS as a class of objects and the frame as the knapsack, the paths in $\mathcal{R}_p$ all constitute a solution to the MCK problem. This proves the *if* part.

Conversely, let $\{x_{1,\alpha_1}, x_{2,\alpha_2}, \ldots, x_{n,\alpha_n}\}$ be a solution to the MCK problem. Then, for each MS$_i$, $i = 1, \ldots, n$, we select a path such that MS$_i$ conserves energy of q_{i,α_i} and the size of

---

[3] Note that the conserved energy of an MS's path is compared to the same MS's path with the most energy cost.

allocated burst(s) to transmit MS$_i$'s data to the BS is u_{i,α_i}. In this way, the conserved energy of all MSs will be $\mathcal{Q}$ and the overall burst size is no larger than $\mathcal{U}$. This constitutes a solution to the ERAD problem, thus proving the *only if* part. □

3. Two heuristics to the ERA problem

Since the ERA problem is NP-complete, finding an optimal solution is impractical due to the time complexity. Thus, we propose two energy-efficient heuristics, the DFA and EFA schemes. Below, we first give the rationale of our heuristics and then depict the DFA and EFA schemes.

3.1. The rationale of our designs

We first observe what the key factors are and how they affect the goal (energy consumption) and the constraint (resource usage) of the ERA problem. Explicitly, we reveal that the transmission rate, the number of concurrent transmissions, and the distance to the receiver (either an RS or the BS) have a great impact on these two terms. To show how these three factors affect the energy consumption and resource usage of each MS, we conduct an experiment as shown in Fig. 3. Consider a network consisting of one BS, four RSs, and four MSs. Each MS selects a distinct RS to relay its data and the network allows four concurrent transmissions. Assume that the distance between each MS and its RS is the same and each MS has an identical uplink demand. Fig. 4 shows the results on normalized energy consumption and resource usage of an MS. In Fig. 4(a), the transmission rate of an MS is normalized by the highest MCS. We can observe that when a lower MCS is used, the MS will need more resources (i.e., frame space) but can reduce its consumed energy. The benefit ratio of the conserved energy to the increased resource usage is more significant when the MS degrades its MCS from a higher level (such as 5 or 6) to a next lower one (such as 4). In this case, the MS can greatly reduce its energy consumption by increasing only a small amount of resource usage. On the other hand, from Fig. 4(b), it can be observed that more concurrent transmissions can decrease resource usage linearly but increase the energy consumption drastically. Although concurrent transmissions can help resource reuse but it harms MSs in terms of the energy consumption. Finally, in Fig. 4(c), it can be observed that the resource usage is not affected by the distance to the receiver when the MCS is fixed, but it can save the consumed energy greatly when the MS chooses a closer RS to relay its data.

From the experiments in Fig. 4, we can obtain two important observations:

- To reduce the energy consumption of an MS, we have to decrease its MCS level (and thus the transmission rate), the number of concurrent transmissions, and the distance to the receiver. However, doing these will also increase the resource usage of the MS. This means that the energy conservation is inversely proportional to the used resource. Thus, we should keep in mind that the
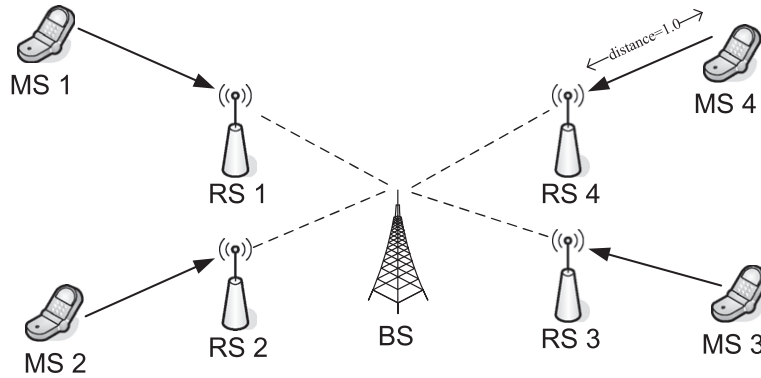
Fig. 3. The example of a transparent-relay network with one BS, four RSs, and four MSs.

overall resource usage of all MSs cannot exceed the frame space constraint when reducing energy.

- The amount of MS's energy reduction is "jointly" decided by its MCS, the number of concurrent transmissions, and the distance to the receiver. In order to minimize the MS's energy consumption, it is insufficient to decrease the three factors individually. Since the experiments show that the benefit ratio of energy decrement to resource increment for each factor is greatly different. An MS may save more energy by considering more than one factor simultaneously. For example, an MS may not be able to relay its data to an RS closer to it because that RS is used by another MS. When considering both the factors of concurrent transmissions and the distance to the receiver, the MS can change to another transmission group and choose such RS to further save energy (even if it may increase the number of concurrent transmissions in that group). This adjusting may be more efficient than that of considering only one factor (such as the MCS). Therefore, we need to consider the possible combination of three factors when trying to reduce MSs' energy consumption.

Based on the two observations and the three key factors, our DFA and EFA heuristics adopt a gradient-like search method to find the suboptimal solutions, as shown in Fig. 5. For ease of presentation, we say that a solution is demand-satisfied if it can satisfy all MSs' demands. Besides, a solution is *feasible* if it is not only demand-satisfied but also the overall frame usage does not exceed the frame space. Given the solution set, DFA first selects a feasible solution which consumes as less frame space as possible to be its starting point. Then, it adopts a forward search to approximate the optimal solution. In each step of search, it tries to adjust the transmission schedule of one MS by evaluating the combinations of the three factors mentioned above such that the new solution is also feasible and the gradient of $\Delta E_D / \Delta S_I$ is maximum, where ΔE_D is the decrement of energy and ΔS_I is the increment of space usage after adjustment. The forward search is repeated until $\Delta E_D / \Delta S_I$ approximates to zero (that is, we cannot further reduce the energy consumption since $\Delta E_D \approx 0$). On the other hand, EFA first selects a demand-satisfied solution that allows MSs to consume as less energy as possible to
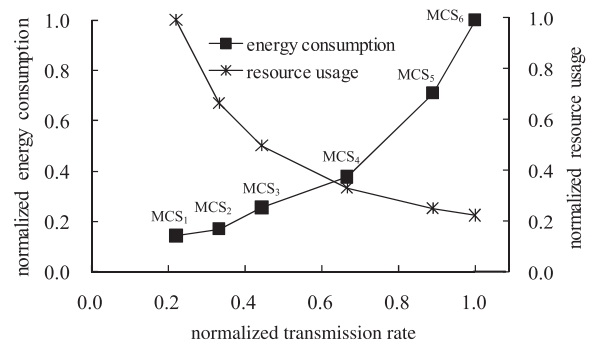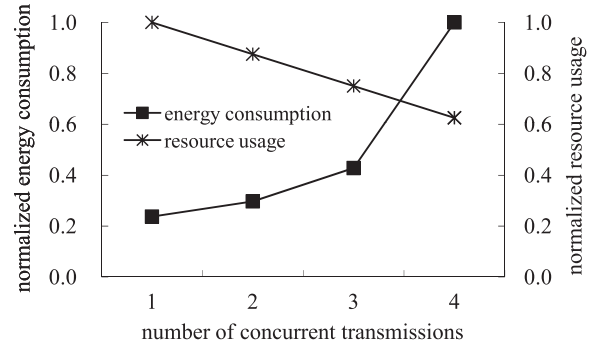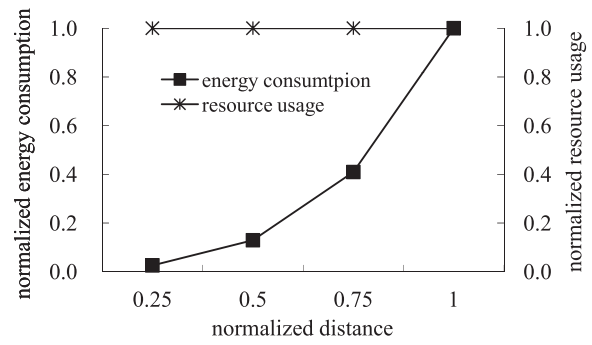
Fig. 4. The effects of rate, concurrent transmission, and receiver distance on energy consumption and resource usage.
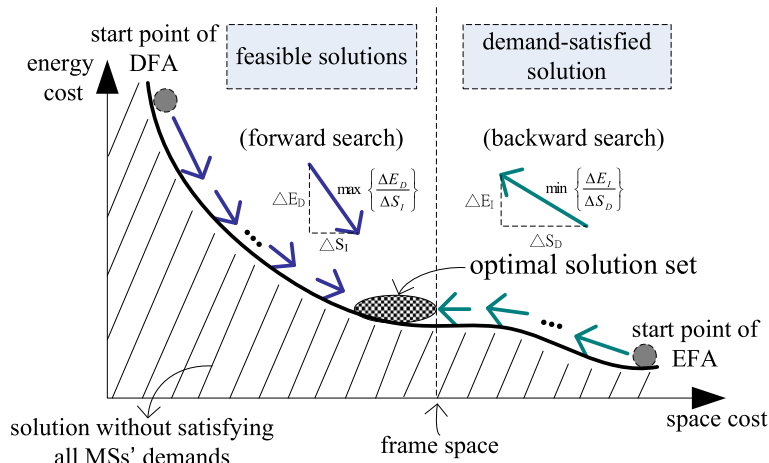
Fig. 5. The concepts of forward and backward searches by the gradient-like method.
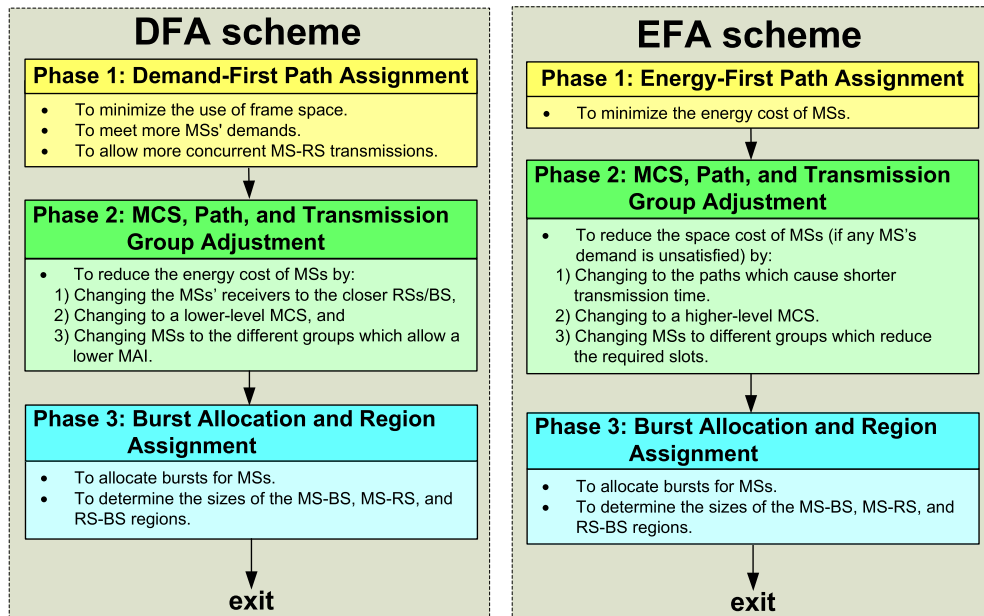

Fig. 6. Flowcharts of our proposed heuristics.

be its starting point. Then, it adopts a backward search to approximate the optimal solution. In each step of search, it tries to adjust the transmission schedule of one MS such that the new schedule is also demand-satisfied while $\Delta E_I/\Delta S_D$ is minimum, where ΔE_I is the increment of energy and ΔS_D is the decrement of space usage after adjustment. The backward search is repeated until the solution becomes feasible.

Fig. 6 shows the flow charts of the two heuristics. In DFA, the first "Demand-First Path Assignment" phase tries to satisfy MSs' demands by selecting the best MCSs and paths and exploiting spatial reuse such that the use of frame space is minimized. However, the above process assumes that each MS transmits at its largest power. So, the

second "MCS, Path, and Transmission Group Adjustment" phase tries to reduce MSs' energy consumption by lowering their transmission rates and adjusting their paths and transmission groups. Each step of reduction is based on the gradient concept. Finally, the third "Burst Allocation and Region Assignment" phase determines the sizes of the MS–BS, MS–RS, and RS–BS regions and allocates uplink bursts for MSs and RSs. On the other hand, EFA first relaxes the frame space constraint to find the initial solution with the minimum total energy consumption in its first "Energy-First Path Assignment" phase. In this phase, MSs choose the closest RSs and the lowest MCSs without spatial reuse. The second "MCS, Path, and Transmission Group Adjustment" phase works based on the gradient concept

to approach the optimum by raising MSs' energy consumption until packing all demands into the frame, i.e., reducing the required space by using more power. The third is the "Burst Allocation and Region Assignment" phase. Since the two schemes start from different initial solutions and apply different strategies, they have different limitations and thus lead to different performances. This will be clear later on.

3.2. Demand-first allocation (DFA) scheme

3.2.1. Phase 1 – burst and path assignment

Assuming that the energy consumption of MSs is not a concern, phase 1 has the following objectives: (i) to minimize the use of frame space, (ii) to meet more MSs' demands, and (iii) to allow more concurrent MS–RS transmissions. This phase helps choose each MS_i's initial path, transmission group, and MCS using the maximum power.

To exploit spatial reuse in the MS–RS region, we model the *maximum allowable interference (MAI)* $\widehat{\mathcal{I}}_{(i,J(i))}^{K(i)}$ at relay $RS_{J(i)}$ if MS_i chooses $RS_{J(i)}$ as its relay using $MCS_{K(i)}$ with power P_i^{MAX}, $i = 1, \ldots, n, J(i) = 0, \ldots, m$, and $K(i) = 1, \ldots, 6$. Recall the $I(i,J(i))$ in Eq. (4), which stands for the current perceived interference for the transmission from MS_i to $RS_{J(i)}$. With the relative distance between MSs and BS/RSs, we can derive the path loss $L(i,J(i))$ of each MS–BS/RS pair. From Eq. (4), each $\widehat{\mathcal{I}}_{(i,J(i))}^{K(i)}$ of an MS_i transmitting to $RS_{J(i)}$ using $MCS_{K(i)}$ with P_i^{MAX} is

$$\widehat{\mathcal{I}}_{(i,J(i))}^{K(i)} = \frac{G_i \cdot G_{J(i)} \cdot P_i^{MAX}}{10^{\frac{\delta(MCS_{K(i)})}{10}} \cdot L(i,J(i))} - B \cdot N_0. \tag{8}$$

We should keep $\widehat{\mathcal{I}}_{(i,J(i))}^{K(i)} \geqslant I(i,J(i))$. Note that using a lower-level MCS can tolerate a higher interference, so $\widehat{\mathcal{I}}_{(i,J(i))}^{K(i)} < \widehat{\mathcal{I}}_{(i,J(i))}^{K(i)-1}$. Also note that for the BS, $\widehat{\mathcal{I}}_{(i,0)}^{K(i)} = 0$, since no concurrent transmission to the BS is allowed. For simplicity, we will pre-calculate all values of $\widehat{\mathcal{I}}_{(i,J(i))}^{K(i)}$ and maintain an MAI table using $(MS_i, RS_{J(i)}, MCS_{K(i)})$ as the index.

Given the network topology $\mathcal{G}$, the path set $\mathcal{R}$, and MS_i's demand d_i, $i = 1, \ldots, n$, phase 1 starts from a set $\mathcal{T}$ of n empty transmission groups and greedily adds more transmission schedules to $\mathcal{T}$, until all frame space is exhausted or all MSs are satisfied. Each transmission schedule has the format $s_i = (RS_{J(i)}, MCS_{K(i)}, P_i)$, which means that MS_i is scheduled to send its data to $RS_{J(i)}$ using $MCS_{K(i)}$ at power P_i. In case that $J(i) \neq 0$, it is implied that $RS_{J(i)}$ will relay MS_i's data to the BS using the best possible MCS level. Note that in this phase, P_i is always equal to P_i^{MAX}.

Step (1) Set all MSs as *unsatisfied*. Set the initial value of $\mathcal{T}$ to be $\{\phi, \phi, \ldots, \phi\}$ (i.e., with n empty sets) and set $F = h \times w$ as the initial amount of free slots.
Step (2) Consider each unsatisfied MS_i. If we adding the path from MS_i to $RS_{J(i)}$ using $MCS_{K(i)}$ to the transmission group $\tau_\ell \in \mathcal{T}$ at power P_i^{MAX} (that is, adding $s_i = \left(RS_{J(i)}, P_i^{MAX}\right)$ to group τ_ℓ), the extra number of slots required will be

$$S_{ex}(s_i, \tau_\ell)$$
$$= \begin{cases} \left\lceil \frac{d_i}{rate(MCS_{K(i)})} \right\rceil, & \text{if } J(i) = 0, \quad \tau_\ell = \phi, \\ \infty, & \text{if } J(i) = 0, \quad \tau_\ell \neq \phi, \\ \max \left\{ \left\lceil \frac{d_i}{rate(MCS_{K(i)})} \right\rceil - S_g(\tau_\ell), 0 \right\} + \left\lceil \frac{d_i}{rate(MCS_{\widehat{K(i)}})} \right\rceil, \\ \qquad \text{if } inf(s_i, \tau_\ell) = TRUE, \quad J(i) \neq 0, \\ \infty, & \text{if } inf(s_i, \tau_\ell) = FALSE, \quad J(i) \neq 0, \end{cases}$$
$$\tag{9}$$

where $S_g(\tau_\ell)$ is the number of slots required by τ_ℓ in the MS–RS region, $inf(s_i, \tau_\ell)$ is a function to determine if adding $s_i = \left(RS_{J(i)}, MCS_{K(i)}, P_i^{MAX}\right)$ to τ_ℓ is interference-free, and $MCS_{\widehat{K(i)}}$ is the best feasible MCS from $RS_{J(i)}$ to the BS. In the first case of $J(i) = 0$, it is the cost to the MS–BS region. In the second case, it means adding an MS–BS transmission to an non-empty group is infeasible. In the third case, it is the extra cost to the MS–RS region plus that to the RS–BS region. In the fourth case, it means adding this path to τ_ℓ is infeasible. Function $inf(s_i, \tau_\ell)$ returns *TRUE* (i.e., interference-free) if and only if the following three conditions are *all* satisfied:

1. $RS_{J(i)}$ does not appear in τ_ℓ. That is, for each $s_a = \left(RS_{J(a)}, MCS_{K(a)}, P_a^{MAX}\right) \in \tau_\ell, J(a) \neq J(i)$.
2. $RS_{J(i)}$ can receive correctly considering all interferences. That is,
$$\sum_{\forall s_a = \left(RS_{J(a)}, MCS_{K(a)}, P_a^{MAX}\right) \in \tau_\ell} \widetilde{P}(a, J(i)) \leqslant \widehat{\mathcal{I}}_{(i,J(i))}^{K(i)}.$$
3. After adding the interference caused by MS_i with $s_i = \left(RS_{J(i)}, MCS_{K(i)}, P_i^{MAX}\right), RS_{J(a)}$ can still receive correctly. That is, for each $s_a = (RS_{J(a)}, MCS_{K(a)}, P_a^{MAX}) \in \tau_\ell, I(a, J(a)) + \widetilde{P}(i, J(a)) \leqslant \widehat{\mathcal{I}}_{(a,J(a))}^{K(a)}$.

   After step (2), we have the extra cost to schedule each unsatisfied MS_i for all combinations of $RS_{J(i)}$, $MCS_{K(i)}$, and τ_ℓ.

Step (3) From the extra costs of all unsatisfied MSs, pick the one causing the least cost of $S_{ex}(s_i, \tau_\ell)$. If $S_{ex}(s_i, \tau_\ell) \leqslant F$, add $s_i = \left(RS_{J(i)}, MCS_{K(i)}, P_i^{MAX}\right)$ to τ_ℓ directly; otherwise, adjust the demand d_i of MS_i proportionally to fit into F and add $s_i = \left(RS_{J(i)}, MCS_{K(i)}, P_i^{MAX}\right)$ to τ_ℓ. Then, update F by deducting the allocated resource and set MS_i as *satisfied*. Also, update $I(a, J(a))$ for each satisfied MS_a in τ_ℓ. Finally, update $S_{ex}(\cdot)$ s of all unsatisfied MSs' schedules for τ_ℓ. Note that after step (3), one MS will be satisfied.
Step (4) If there still is space in an uplink subframe and there still exists any unsatisfied MS, go back to step (3); otherwise, go to the next phase.

3.2.2. Phase 2 – MCS, path, and group adjustment

Phase 1 aims at reducing the use of frame space, but the maximum powers have been used by all MSs. This phase

tries to make some adjustments and lower their energy costs by taking advantage of the extra free frame space F. We try three possibilities to reduce an MS's energy: (i) Change its receiver to a closer RS/BS. (ii) Change to a lower-level MCS. (iii) Change to a different transmission group with a different MCS and receiver. In particular, for possibility (ii), recall that the energy cost of MS_i can be written as $E_i = T_i \times P_i = \left\lceil \frac{d_i}{rate(MCS_{K(i)})} \right\rceil \times P_i$. By ignoring the ceiling function and assuming a fixed interference level of $B \cdot N_o + I(i, J(i))$, the energy cost per bit to reach the SINR in Table 1 can be written as

$$E_i = \frac{1}{rate(MCS_{K(i)})} \times (10^{\frac{\delta(MCS_{K(i)})}{10}} \beta), \tag{10}$$

where $\beta = \frac{(B \cdot N_o + I(i, J(i))) \cdot L(i, J(i))}{G_i \cdot G_{J(i)}} > 0$. In Table 2, we do see that the energy cost per bit decreases as the MCS level decreases.

Given the current set $\mathcal{T}$ and the remaining free resource F from phase 1, phase 2 works as follows:

Step (1) For each $\tau_\ell \in \mathcal{T}$, consider each transmission schedule $s_i = (RS_{J(i)}, MCS_{K(i)}, P_i) \in \tau_\ell$. There are three possibilities for MS_i to reduce its energy: (a) Change its MCS and power. (b) Change its relay and power. (c) Change its group, relay, MCS, and power. For each s_i, we may find multiple combinations of $s_i' = (RS_{J'(i)}, MCS_{K'(i)}, P_i')$ and $\tau_{\ell'}$ such that s_i' is the new transmission schedule for MS_i and $\tau_{\ell'}$ is the transmission group to accommodate s_i' (which may or may not be equal to τ_ℓ).
To find all feasible s_i' and $\tau_{\ell'}$, let us consider the above three cases. In case (a), since $RS_{J(i)}$ is unchanged, we can simply try different $MCS_{K'(i)}$ and then use Eq. (4) based on the existing interference $I(i, J(i))$ perceived by $RS_{J(i)}$ to compute the best power P_i'. With this new power P_i', we also need to check if this would exceed the tolerable interference of any other RS in τ_ℓ. If so, this transmission schedule is not feasible. In case (b), since τ_ℓ is unchanged, we try other unused RSs in τ_ℓ and follow the procedure in case (a) to find appropriate MCSs and power. Similarly, we need to check if this would excess interference to existing RSs. In case (c), we will try to delete s_i from τ_ℓ and add MS_i's demand to other $\tau_{\ell'}$. For each $\tau_{\ell'}$, the same procedure in case (b) can be used to identify all possible s_i'.
Note that after step (1), we have all new feasible s_i' and $\tau_{\ell'}$ for MS_i.

Step (2) For each $(s_i', \tau_{\ell'})$ pair, we calculate the saving of energy and the cost of extra slots for MS_i to make this change. The saving of energy is written as

$$\Delta_E \big((s_i, \tau_\ell), (s_i', \tau_{\ell'}) \big) = \left(\left\lceil \frac{d_i}{rate(MCS_{K(i)})} \right\rceil \times P_i \right)$$
$$- \left(\left\lceil \frac{d_i}{rate(MCS_{K'(i)})} \right\rceil \times P_i' \right). \tag{11}$$

Then, the cost of extra slots is derived as

$$\Delta_S \big((s_i, \tau_\ell), (s_i', \tau_{\ell'}) \big)$$
$$= \begin{cases} \big(S_{tot}(\tau_\ell - \{s_i\}) + S_{tot}(\tau_{\ell'} \cup \{s_i'\}) \big) \\ - (S_{tot}(\tau_\ell) + S_{tot}(\tau_{\ell'})), & \text{if } \tau_\ell \neq \tau_{\ell'}, \\ S_{tot}(\tau_{\ell'} - \{s_i\} \cup \{s_i'\}) - S_{tot}(\tau_\ell), & \text{if } \tau_\ell = \tau_{\ell'}, \end{cases} \tag{12}$$

Note that $\Delta_S \big((s_i, \tau_\ell), (s_i', \tau_{\ell'}) \big)$ should not exceed the available resource F and the saving $\Delta_E \big((s_i, \tau_\ell), (s_i', \tau_{\ell'}) \big)$ should be positive. Otherwise, this pair $(s_i', \tau_{\ell'})$ is infeasible and should not be considered.

Step (3) From all feasible pairs $(s_i', \tau_{\ell'})$, we use the *energy-per-extra-slot ratio*

$$\frac{\Delta_E \big((s_i, \tau_\ell), (s_i', \tau_{\ell'}) \big)}{\Delta_S \big((s_i, \tau_\ell), (s_i', \tau_{\ell'}) \big)}$$

as the metric (this is recognized as the "gradient" in our scheme). The $(s_i', \tau_{\ell'})$ pair with the largest ratio is selected (this represents the "steepest gradient" in the energy cost). Then, we remove s_i from τ_ℓ, add s_i' to $\tau_{\ell'}$, deduct $\Delta_S((s_i, \tau_\ell), (s_i', \tau_{\ell'}))$ from F, and update all interference levels of all RSs in τ_ℓ and $\tau_{\ell'}$. Then, we calculate $\Delta_E(\cdot)$ and $\Delta_S(\cdot)$ for each schedule s_a in τ_ℓ and each schedule s_b in $\tau_{\ell'}$. If any change in τ_ℓ and $\tau_{\ell'}$ is done, go to step (3); otherwise, go to the next phase.

We make some remarks below. First, updating an MS's power level is possible even if no extra slots are needed. The reason is that when an MS lowers its power, other RSs may experience lower interference levels, making it possible for other MSs to meet the required SINRs using lower power. From our experience, such a positive cycle would repeatedly benefit lots of MSs. Second, the above process will eventually terminate. To speed up our algorithm, we can set a threshold ∂ on Δ_E or on the number of iterations.

3.2.3. Phase 3 – burst allocation and region assignment

After phase 2, all MSs' paths, MCSs, power, and transmission groups are determined. This phase will allocate bursts for MSs and determine the sizes of the MS–BS, MS–RS, and RS–BS regions accordingly.

Given the current set $\mathcal{T}$ from Phase 2, Phase 3 works as follows.

Table 2
Energy costs per bit for different MCSs.

Level k	Energy cost (mW)
1	0.082β
2	0.098β
3	0.147β
4	0.219β
5	0.413β
6	0.582β

Step (1) Let $R_{MS--BS}(\mathcal{T}), R_{MS--RS}(\mathcal{T})$, and $R_{RS--BS}(\mathcal{T})$ be the sizes of the MS–BS, MS–RS, and RS–BS regions, respectively. Calculate them as follows:

$$R_{MS--BS}(\mathcal{T}) = \sum_{\forall \tau_\ell \in \mathcal{T}} \sum_{\forall s_i = (RS_{J(i)}, MCS_{K(i)}, P_i) \in \tau_\ell: J(i)=0} \left\lceil \frac{d_i}{rate(MCS_{K(i)})} \right\rceil,$$

$$R_{MS--RS}(\mathcal{T}) = \sum_{\forall \tau_\ell \in \mathcal{T}: s_i = (RS_{J(i)}, MCS_{K(i)}, P_i) \in \tau_\ell, J(i) \neq 0} S_g(\tau_\ell),$$

$$R_{RS--BS}(\mathcal{T}) = \sum_{\forall \tau_\ell \in \mathcal{T}} \sum_{\forall s_i = (RS_{J(i)}, MCS_{K(i)}, P_i) \in \tau_\ell, J(i) \neq 0} \left\lceil \frac{d_i}{rate(MCS_{\widehat{K}(i)})} \right\rceil.$$

$$\tag{13}$$

Step (2) According to each schedule $s_i = (RS_{J(i)}, MCS_{K(i)}, P_i)$ in $\tau_\ell \in \mathcal{T}$, allocate MS_i the corresponding burst(s) to the MS–BS, MS–RS, and RS–BS regions accordingly.

To summarize, the DFA scheme finds its best solution by first calculating a temporal solution that can consume the minimum frame space and then iteratively refines the solution to reduce MSs' energy consumption. The above refinement is repeated until either the frame space is exhausted or the total energy consumption is minimized. However, deriving the minimal space solution (in phase 1) takes a lot of time. In addition, phase 2 might face convergence problem because the value of energy-per-extra-slot ratio is usually difficult to converge since each MCS, path, and group adjustment in phase 2 may incur a chain reaction such that a large number of iterations will be required to reach its best solution. Therefore, we apply a threshold to limit the number of iterations in phase 2 to guarantee the convergence of DFA. In the next section, we will discuss how to address the convergence issue.

3.3. Energy-first allocation (EFA) scheme

To solve the problem in DFA, EFA makes the following improvements:

1. EFA first relaxes the constraint of frame space so that it can easily find a temporal solution which consumes the least energy as the starting point. This significantly reduces the computational complexity.
2. Unlike DFA that reduces the energy consumption (which is continuous) in phase 2, EFA tries to reduce the frame usage in a discrete manner (because the basic unit of the frame space is a slot). This not only alleviates the computation cost but also guarantees the convergence of EFA.
3. EFA adopts simultaneous equations to calculate the minimum transmission power of MSs in each transmission group. This can help to further reduce the energy consumption of MSs.

The EFA scheme starts with a trivial set $\mathcal{T}$ of transmission groups where each group contains only one MS with the closest RS/BS using the lowest MCS. It is thus a solution with the least energy cost. However, the total number of slots required may exceed the frame space. We then adjust these schedules by changing their power, MCSs, paths, and

transmission groups based on gradient-like search, until they fit into one frame space.

Phase 1: For each MS_i, we create a transmission schedule $s_i = (RS_{J(i)}, MCS_{K(i)}, P_i)$ such that $RS_{J(i)}$ is the closest to MS_i, $MCS_{K(i)} = MCS_6$ (the lowest one), and P_i is the lowest power required to communicate with $RS_{J(i)}$. Then, we let each s_i be in one transmission group by setting $\tau_i = \{s_i\}$, $i = 1, \ldots, n$. Let L be the total required slots of $\mathcal{T}$. Initially, $L = R_{MS--BS}(\mathcal{T}) + R_{MS--RS}(\mathcal{T}) + R_{RS--BS}(\mathcal{T})$. Finally, check whether $L \leqslant w \times h$. If yes, go to phase 3. Otherwise, go to phase 2 to reduce the space cost for possibilities.

Phase 2: For each $\tau_\ell \in \mathcal{T}$, consider the transmission schedule $s_i = (RS_{J(i)}, MCS_{K(i)}, P_i) \in \tau_\ell$. There are three possibilities for MS_i to reduce the space cost. (a) Within the same group τ_ℓ, MS_i can still transmit to $RS_{J(i)}$ but using a higher MCS. (b) Within the same group τ_ℓ, MS_i can still use $MCS_{K(i)}$ but changing its relay. (Note that the best feasible MCS for each RS to the BS may be different so that the space cost will be also different). (c) MS_i switches to another group and then selects proper MCS and relay. For each possibility, we use s_i' as the new schedule for MS_i and $\tau_{\ell'}$ as the new group accommodating s_i'.

Step (1) To find all feasible s_i' and $\tau_{\ell'}$, we consider the above possibilities (a), (b), and (c). Unlike DFA, EFA tries to further reduce energy by optimizing the transmission power of multiple MSs in the transmission group $\tau_{\ell'}$ when s_i' joins it. Therefore, we propose using simultaneous equations to derive the minimum required power of all MSs in group $\tau_{\ell'}$. Suppose that if adding s_i' to $\tau_{\ell'}$, we have a set of schedules $\{s_a = (RS_{J(a)}, MCS_{K(a)}, P_a)\}$ in $\tau_{\ell'}, |\tau_{\ell'}| = z$, where $J(a)$ and $K(a)$ are the indexes of the RS and MCS used by MS_a. Let P_a be the power of MS_a, $0 \leqslant P_a \leqslant P_a^{MAX}$. It follows that the SINR perceived by $RS_{J(a)}$ should be over $\delta(MCS_{K(a)})$, i.e.,

$$SINR(a, J(a)) = 10 \cdot \log_{10} \left(\frac{\widetilde{P}(a, J(a))}{B \cdot N_o + I(a, J(a))} \right)$$
$$\geqslant \delta(MCS_{K(a)}). \tag{14}$$

To minimize the power, we make the equal mark (i.e., "=") hold. Thus, we have

$$\frac{\frac{G_a \cdot G_{J(a)} \cdot P_a}{L(a, J(a))}}{B \cdot N_o + \sum_{s_{a'} \in \tau_{\ell'}, a' \neq a} \frac{G_{a'} \cdot G_{J(a)} \cdot P_{a'}}{L(a', J(a))}} = 10^{\frac{\delta(MCS_{K(a)})}{10}}. \tag{15}$$

Since the right-hand side is a constant, Eq. (15) can be converted into a simultaneous equations for each $p_a, s_a \in \tau_{\ell'}$. Repeating this for each $MS_a, s_a \in \tau_{\ell'}$, we obtain z equalities. Then, by solving these equalities, we can find the best power P_a for each MS_a in $\tau_{\ell'}$ in polynomial time and check whether they are feasible for

concurrent transmissions by $P_a \leqslant P_a^{MAX}$.

After step (1), we have all new feasible s_i' and $\tau_{\ell'}$ for MS_i.

Step (2) For each $(s_i', \tau_{\ell'})$ pair, we calculate the cost of extra consumed energy and saving of slots for MS_i to make this change. Given any transmission group τ, let $E_g(\tau)$ be the summation of energy consumed by all transmission schedule $s_a = (RS_{J(a)}, MCS_{K(a)}, P_a)$ in τ, which can be defined as

$$E_g(\tau) = \sum_{s_a=(RS_{J(a)},MCS_{K(a)},P_a)\in\tau} \left\lceil \frac{d_a}{rate(MCS_{K(a)})} \right\rceil \times P_a.$$

(16)

Then, the cost of extra consumed energy is derived as

$$\Delta_\Psi\big((s_i,\tau_\ell),(s_i',\tau_{\ell'})\big)$$
$$= \begin{cases} (E_g(\tau_\ell - \{s_i\}) + E_g(\tau_{\ell'} \cup \{s_i'\})) \\ -(E_g(\tau_\ell) + E_g(\tau_{\ell'})), & \text{if } \tau_\ell \neq \tau_{\ell'}, \\ E_g(\tau_{\ell'} - \{s_i\} \cup \{s_i'\}) - E_g(\tau_\ell), & \text{if } \tau_\ell = \tau_{\ell'}. \end{cases}$$

(17)

The saving of slots is written as

$$\Delta_\Omega\big((s_i,\tau_\ell),(s_i',\tau_{\ell'})\big)$$
$$= \begin{cases} (S_{tot}(\tau_\ell) + S_{tot}(\tau_{\ell'})) \\ -(S_{tot}(\tau_\ell - \{s_i\}) + S_{tot}(\tau_{\ell'} \cup \{s_i'\})), & \text{if } \tau_\ell \neq \tau_{\ell'}, \\ S_{tot}(\tau_\ell) - S_{tot}(\tau_{\ell'} - \{s_i\} \cup \{s_i'\}), & \text{if } \tau_\ell = \tau_{\ell'}. \end{cases}$$

(18)

Note that $\Delta_\Omega\big((s_i,\tau_\ell),(s_i',\tau_{\ell'})\big)$ should be positive. Otherwise, this pair $(s_i',\tau_{\ell'})$ provides no benefit and should not be considered.

Step (3) From all feasible pairs $(s_i',\tau_{\ell'})$, we use the *slot-per-extra-energy ratio*

$$\frac{\Delta_\Omega\big((s_i,\tau_\ell),(s_i',\tau_{\ell'})\big)}{\Delta_\Psi\big((s_i,\tau_\ell),(s_i',\tau_{\ell'})\big)},$$

as the metric (this is recognized the as "gradient" in our scheme). The $(s_i',\tau_{\ell'})$ pair with the largest ratio is selected (which represents the "steepest gradient" in space cost). Then, we remove s_i from τ_ℓ, add s_i' to $\tau_{\ell'}$, deduct $\Delta_\Omega\big((s_i,\tau_\ell),(s_i',\tau_{\ell'})\big)$ from L. Then, we recalculate $\Delta_\Psi(\cdot)$ and $\Delta_\Omega(\cdot)$ for each schedule s_a in τ_ℓ and each schedule s_b in $\tau_{\ell'}$ accordingly. Go back to step (3) if $L > w \times h$ and there is any change in τ_ℓ or $\tau_{\ell'}$; otherwise, go to the next phase.

Note that it is possible that more than two schedules have the largest burst size in a group. By changing one of them, no space is saved. In this case, we can try to raise their MCSs by one level simultaneously to further reduce the space cost.

Phase 3: If the total number of required slots still exceeds the frame space, i.e., $L > w \times h$, we can shrink the sizes of some MSs' bursts until the overall allocation can fit the frame space. Then, we adopt the phase 3 in DFA to allocate

bursts and determine the sizes of the MS–BS, MS–RS, and RS–BS regions accordingly.

3.4. Analysis of time complexity

For DFA, phase 1 initially costs $O(n(m+1)6) = O(nm)$ to model the MAI for all MSs transmitting to all possible RSs/BS with six MCSs, where $(m+1)$ means that there are m relay paths and one direct path to the BS. In step (1), it costs $O(n)$. In step (2), for each schedule, it costs $O(m)$ because it has at most m schedules in a transmission group to be verified whether adding the new schedule is interference-free. Then, since we may have at most $O(n(m+1)6)$ possible schedules and n possible transmission groups for all MSs, the time complexity of step (2) is $O(m) \cdot O(n(m+1)6) \cdot n = O(n^2m^2)$. In step (3), it costs $O(n^2m^2)$ because it has at most $(n(m+1)6) \cdot n$ schedules to be picked and at most n schedules to be updated for their costs to that group (each costs $O(m^2)$). In step (4), it may go back to step (3) at most n times since there are n MSs. Therefore, the time complexity of phase 1 in the DFA scheme costs $O(n) + O(n^2m^2) + n \cdot O(n^2m) = O(n^3m^2)$. For phase 2, step (1) and 2 cost $O(n^2m^2)$ because there are at most $(n(m+1)6)$ possible new schedules and n possible groups to be tried. Then, each schedule needs to verify whether they are interference-free (which costs $O(m)$). Thus, it can calculate the extra cost and conserved energy accordingly. In step (3), it costs $O(n^2m^2)$ because it has at most $(n(m+1)6) \cdot n$ schedules to be chosen and at most n schedules to be updated for their costs to those groups (each costs $O(m^2)$). Besides, it will go back to step (3) at most $\frac{P_i^{MAX} \cdot (w \cdot h)}{\partial}$ times, where $P_i^{MAX} \cdot (w \cdot h)$ is the maximum energy cost of an MS and ∂ is a threshold on Δ_E. Since we have n MSs, phase 2 costs $n \cdot \frac{P_i^{MAX} \cdot (w \cdot h)}{\partial} \cdot O(n^2m^2) = O(n^3m^2)$ if n and m are sufficiently large. For phase 3, it costs $O(n)$ to calculate the region sizes and to allocate at most $2n$ bursts if all MSs use relays to transmit data. Therefore, the DFA scheme costs $O(n^3m^2) + O(n^3m^2) + O(n) = (n^3m^2)$.

For EFA, in phase 1, it costs $O(nm)$ to choose the closest RS and lowest MCS for each MS. In phase 2, step (1) costs $O(n)$. In step (2), each schedule costs $O(m^3)$ to solve m simultaneous equations by the *Gaussian Elimination* because there are at most m transmission schedules in one transmission group. Since we may have at most $O(n(m+1)6)$ possible schedules and n possible transmission groups, the time complexity of step (2) is $O(n^2m^4)$. In step (3), it costs $O(n^2m^4)$ because it has no more than $(n(m+1)6) \cdot n$ schedules to be chosen and then takes $O(nm^4)$ to update. Besides, it will go back to step (1) at most $L - (w \cdot h)$ times, where L is the total number of slots required by the schedules in phase 1, which is proportional to the number of demands (i.e., n). Phase 3 costs $O(n)$. Therefore, EFA scheme costs $O(nm) + \{O(n) + O(n^2m^4) + [L - (w \cdot h)] \cdot O(n^2m^4)\} + O(n) = O(n^3m^4)$.

4. Performance evaluation

In this section, we develop a simulator in Java to verify the effectiveness of our heuristics. The system parameters

Table 3
The parameters in our simulator.

Parameter	Value
Channel bandwidth	10 MHz
FFT size	1024
Zone category	PUSC with reuse 1
Slot-time	200.94 µs
Uplink frame duration	2.5 ms
Uplink subframe space	12×30
MCS	Table 1
Traffic	UGS, rtPS, nrtPS, and BE
Demand d_i	Table 4
Path loss model	two-ray ground
Thermal noise	−100 dBm
P_i^{MAX}	1000 mW (milliwatt)
Threshold ∂	50

Table 4
The traffic model used in our simulator.

| Traffic class | Traffic type | Bandwidth (bytes/frame) | | |
		Minimum	Maximum	Average
UGS	CBR	40–150	50–150	50–150
rtPS	VBR	50–100	100–150	75–125
	Gaming	1.2	3	1.2
	VoIP	1.4	1.4	1.4
nrtPS	VBR	50–100	100–125	75–125
	FTP	4	10	7
	Real trace	40	110	85
BE	VBR	0	0–150	0–75
	HTTP	0	7	3.6

of our simulator are listed in Table 3. We consider four types of traffic: UGS, rtPS, nrtPS, and BE. Table 4 lists the parameters used to model these traffic. The network contains one BS and several RSs and MSs. RSs are uniformly deployed inside the 2/3 coverage range of the BS to get the best performance gain [22,23] and the number of RSs is ranged from 0 to 32. MSs are randomly deployed inside the BS's coverage and the number of MSs is ranged from 10 to 80. Each MS may move inside the BS's coverage following the random waypoint model with the maximal speed of 20 meters per second [31].

We compare our proposed DFA and EFA schemes against the *minimal-coloring (MC) scheme* [22,23] and the *modified solution of MCK problem (sMCKP)* [24]. The MC scheme considers spatial reuse while the sMCKP scheme addresses the energy consumption of MSs. Specifically, the MC scheme first selects a path with the minimum transmission time (by using the highest MCS level) for each MS. Then, this scheme assigns one color for those MS–RS communications that can coexist and tries to use the minimum number of colors. In this way, the spatial reuse can be realized. On the other hand, the sMCKP scheme calculates a benefit value of each MS, which is defined by the ratio of the amount of energy reduction to the increase of burst size when the MS changes from its current MCS level to another level. Then, sMCKP iteratively selects one MS with the maximum benefit value and changes its MCS accordingly, until the maximum benefit is zero. However, sMCKP does not exploit RSs to help relay MSs' data.

For the MC scheme and our heuristics, we use the terms "-SR" and "-NSR" to indicate whether or not they adopt spatial reuse. In our heuristics, we can set the MAI values as zeros for the DFA scheme and keep the schedules in original groups for the EFA scheme to realize *no* spatial reuse.

In addition, to further investigate the performance of our proposed schemes. We define two ideal performance boundaries in terms of *energy consumption lower bound (ELB)* and *demand satisfaction ratio upper bound (DUB)*. ELB assigns each MS a schedule in a group containing only itself and chooses a closest RS/BS as its receiver using the lowest MCS without consideration of frame space limitation. ELB is expressed as follows.

$$\sum_{i=1,\ldots,n} \left\lceil \frac{d_i}{rate(MCS_6)} \right\rceil \times \frac{10^{\frac{\delta(MCS_6)}{10}}(B \cdot N_o + 0) \cdot L(i,j^*)}{G_i \cdot G_{j^*}}, \qquad (19)$$

where $j^* = \mathrm{argmin}_{J(i)=1,\ldots,m}\{L(i,J(i))\}$. The right part of Eq. (19) is the transmission power derived from Eq. (4) by adopting the equal sign. On the other hand, DUB schedules each MS to transmit to the BS if its required slots is less than that the MS's RS required to transmit to the BS. In addition, we assume that each transmission group can accommodate the number of MS–RS transmissions up to the number of RSs in the network, i.e., DUB considers the interference perceived at any RS as zero no matter there are concurrent MS–RS transmissions or not (thus so called *ideal*). Hence, DUB can be expressed as $\min\{\frac{F}{L}, 1\}$, where L is the total required slots, defined by

$$L = \sum_{i \in \mathcal{I}} \left\lceil \frac{d_i}{rate(MCS_{K^B(i)})} \right\rceil + \sum_{i \notin \mathcal{I}} \frac{\left\lceil \frac{d_i}{rate(MCS_{K^R(i)})} \right\rceil}{m}$$

$$+ \sum_{i \notin \mathcal{I}} \left\lceil \frac{d_i}{rate(MCS_{\widehat{K}(i)})} \right\rceil. \qquad (20)$$

$\mathcal{I}$ is a set of MSs with the BS as its receiver, i.e.,

$$\mathcal{I} = \left\{ i \middle| \left\lceil \frac{d_i}{rate(MCS_{K^B(i)})} \right\rceil < 0 + \left\lceil \frac{d_i}{rate(MCS_{\widehat{K}(i)})} \right\rceil, \ i = 1,\ldots,n \right\},$$

and $MCS_{K^B(i)}$ and $MCS_{K^R(i)}$ are the highest feasible MCSs of MS_i transmitting to the BS and the RS, respectively. The first part of Eq. (20) is the number of required slots in MS–BS region. The second part and the third part of Eq. (20) are the costs in MS–RS and RS–BS regions, respectively. Now, let's explain why DUB takes the MS–RS cost as the second part of Eq. (20). As we know, the number of required slots of an MS–RS transmissions is determined by the largest burst size in all MS–RS transmissions of the corresponding transmission group. Assume we have G non-empty transmission groups in the MS–RS region, τ_ℓ, $\ell = 1,\ldots,G$. Let V_ℓ, $\ell = 1,\ldots,G$, be the largest burst size in the ℓth transmission group. It is known that the following equation is established,

$$\sum_{i \notin \mathcal{I}} \left\lceil \frac{d_i}{rate(MCS_{K^R(i)})} \right\rceil = \sum_{\ell=1,\dots,G} \sum_{s_i \in \tau_\ell} \left\lceil \frac{d_i}{rate(MCS_{K^R(i)})} \right\rceil$$

$$\leqslant \sum_{\ell=1,\dots,G} |\tau_\ell| \cdot V_\ell.$$

From above equation, we can derive that

$$\sum_{\ell=1,\dots,G} V_\ell \geqslant \sum_{i \notin \mathcal{I}} \frac{\left\lceil \frac{d_i}{rate(MCS_{K^R(i)})} \right\rceil}{|\tau_\ell|} \geqslant \sum_{i \notin \mathcal{I}} \frac{\left\lceil \frac{d_i}{rate(MCS_{K^R(i)})} \right\rceil}{m}.$$

Hence, the second part of Eq. (20) is a lower bound of the number of required slots in the MS–RS region.

4.1. Energy consumption

We first evaluate the total energy consumption of MSs per frame under different numbers of MSs, as shown in Fig. 7. The number of RSs is 8 and the network is under the non-saturated condition. Note that the y-axis is drawn with exponential scales. Clearly, the energy consumption of MSs under all schemes increases when the number of MSs increases. The sMCKP scheme makes MSs consume the most energy because it does not exploit RSs to reduce the transmission power of MSs. For the case without spatial reuse, the proposed DFA-NSR and EFA-NSR schemes can save energy up to 72% and 80% of MSs' energy, respectively, compared with the MC-NSR scheme. The reason is that the proposed schemes can determine better MCSs and closer RSs for MSs to conserve energy. On the other hand, by allowing spatial reuse, the proposed DFA-SR and EFA-SR schemes can reduce unnecessary energy consumption of MSs compared to the ones without spatial reuse. Although the MC-SR scheme adopts spatial reuse to allow concurrent transmissions, it does not change MSs' paths or lower MCSs for energy conservation when the free resource remains. Thus, it outperforms the case without spatial reuse. In addition, we can observe that EFA-SR scheme saves more energy than the DFA scheme. This is because EFA scheme exploits the optimal power, deriving by the simultaneous equations, when conducting spatial reuse. Fig. 7 shows that the proposed DFA-SR and EFA-SR schemes can save up to 86% and 92% of MSs' energy,
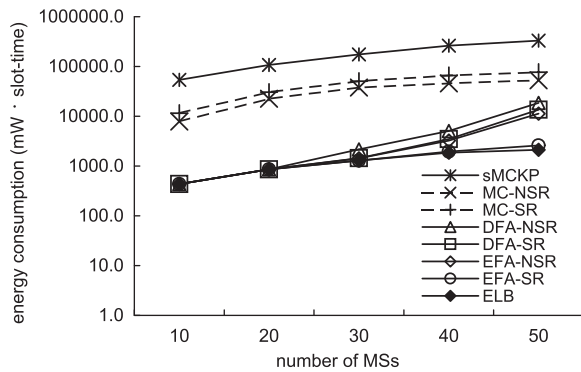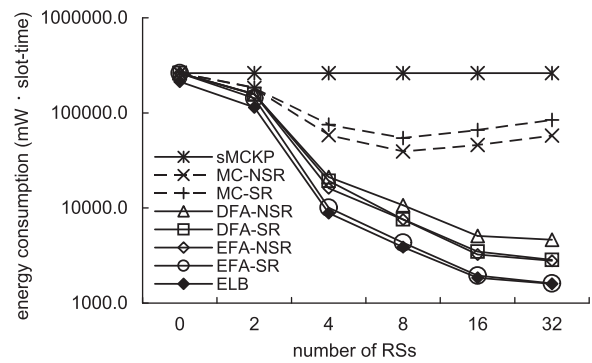

Fig. 8. The energy consumption of MSs under different numbers of RSs, where there are 50 MSs.

respectively, compared with the MC-SR scheme. It is important to note that the performance of our EFA-SR scheme approximates to the energy consumption lower bound. Specifically, when the number of MSs is 10, 20, 30, 40, and 50, the performance errors between the DFA-SR/EFA-SR schemes and the energy consumption lower bound are 0%/0%, 0%/0%, 22%/0.2%, 95%/11.0%, and 589%/39.0%, respectively.

We then measure the total energy consumption of MSs under different numbers of RSs, as shown in Fig. 8. Note that the y-axis is drawn with exponential scales. Since the sMCKP scheme does not exploit RSs, its energy consumption is always the same. On the other hand, the energy consumption of the MC scheme and our heuristics decreases when the number of RSs increases because each MS has more RSs to select to save its energy. Similarly, for the case without spatial reuse, the DFA-NSR and EFA-NSR schemes can save energy up to 77% and 85% of MSs' energy, respectively, compared with the MC-NSR scheme. Furthermore, by allowing spatial reuse, the proposed DFA-SR and EFA-SR schemes outperform other schemes. From Fig. 8, the proposed DFA-SR and EFA-SR schemes can save up to 90% and 98% of MSs' energy, respectively, compared with the MC-SR scheme. It is important to note that the performance of our EFA-SR scheme approximates to the energy consumption lower bound. Specifically, when the number of RSs is 0, 2, 4, 8, 16, and 32, the performance errors between the DFA-SR/EFA-SR schemes and the energy consumption lower bound are 21%/21%, 61%/56%, 490%/56%, 589%/39%, 539%/22%, and 485%/16%, respectively.

4.2. Satisfaction ratio

Next, we investigate the *satisfaction ratio* of MSs, which is defined by the ratio of the amount of *satisfied* demands to the total amount of demands per frame. When the satisfaction ratio is 1, it means that the scheme can satisfy all MSs' demands. Fig. 9 shows the satisfaction ratios of all schemes under different numbers of MSs, where the number of RSs is 32. When there are less than 30 MSs, all schemes have a satisfaction ratio of 1 because the network is not saturated. The sMCKP scheme has the lowest satisfaction ratio when the number of MSs is more than 30, because this scheme does not exploit RSs to improve network capacity. Without
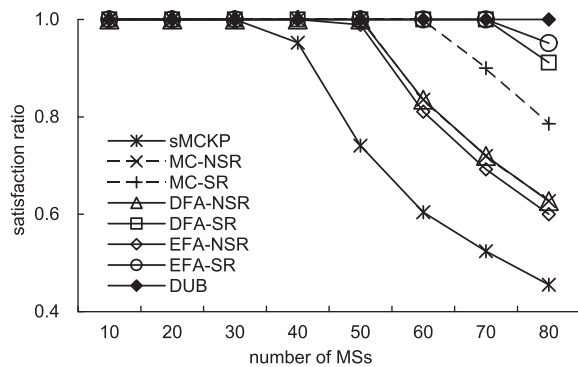

Fig. 7. The energy consumption of MSs under different numbers of MSs, where there are 8 RSs.

Fig. 9. The satisfaction ratio of MSs under different numbers of MSs, where there are 32 RSs.
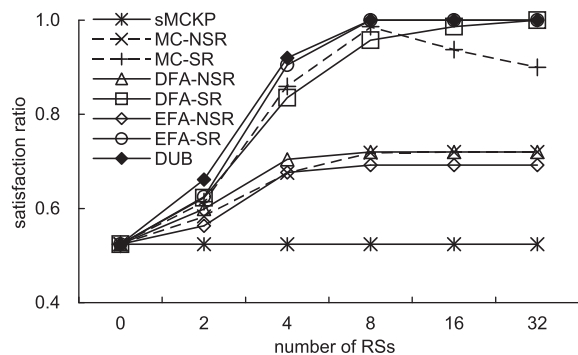

Fig. 10. The satisfaction ratio of MSs under different numbers of RSs, where there are 70 MSs.

spatial reuse, the satisfaction ratios of the MC-NSR scheme and the proposed heuristics, DFA-NSR and EFA-NSR schemes, are similar. However, by exploiting spatial reuse, the proposed schemes always have higher satisfaction ratios than other schemes. The EFA-SR scheme performs the best because it can compactly overlap bursts to satisfy more MSs' demands. It is important to note that the performance of our EFA-SR scheme approximates to the demand satisfaction ratio upper bound. Specifically, when the number of MSs is 10, 20, 30, 40, and 50, the performance errors between the DFA-SR/EFA-SR schemes and the demand satisfaction ratio upper bound are 0%/0%, 0%/0%, 0%/0%, 0%/0%, and 6%/5%, respectively.

Fig. 10 shows the satisfaction ratios of all schemes under different numbers of RSs, where the number of MSs is 70. Again, the satisfaction ratio of the sMCKP scheme is not affected by the number of RSs because it does not consider the existence of RSs. Without spatial reuse, our heuristics, the DFA-NSR and EFA-NSR schemes, perform similarly to the MC-NSR scheme. With spatial reuse, when the number of RSs is more than 8, increasing the number of RSs will decrease the satisfaction ratio of the MC-SR scheme. The reason is that the MC-SR scheme makes all MSs transmit at their highest MCS levels. In this case, more interference may arise when there are more RSs. On the other hand, the proposed DFA-SR and EFA-SR schemes can better utilize RSs than the MC-SR scheme such that

the satisfaction ratio increases when the number of RSs increases. Especially when the number of RSs is more than 4, the EFA-SR scheme always has a satisfaction ratio of 1. This is because it uses RSs to fully exploit spatial reuse and compactly overlap bursts to satisfy the demands of MSs. It is important to note that the performance of our EFA-SR scheme approximates the demand satisfaction ratio upper bound. Specifically, when the number of RSs is 0, 2, 4, 8, 16, and 32, the performance errors between the DFA-SR/EFA-SR schemes and the demand satisfaction ratio upper bound are 0%/0%, 6%/5%, 9%/2%, 4%/0%, 1%/0%, and 0%/0%, respectively.

5. Conclusions

In this paper, we have addressed the energy conservation issue in the uplink resource allocation problem of an 802.16j transparent-relay network. We show this problem to be NP-complete and point out that using a higher MCS level and allowing more concurrent transmissions may harm an MS in terms of its energy consumption. We have proposed two energy-efficient heuristics with the different allocation strategies. The key idea is that we determine the better MCSs, paths, and transmission groups to adjust the use of the frame space and thus to satisfy more MSs' demands while reduce their energy consumption. Simulation results have verified the effectiveness of our heuristics, where our heuristics can save more energy of MSs while increasing their satisfaction ratios as compared with existing schemes.

Acknowledgment

References

[1] IEEE Standard 802.16-2009, IEEE standard for local and metropolitan area networks part 16: Air interface for fixed broadband wireless access systems, 2009.

[2] IEEE Standard 802.16j-2009, IEEE standard for local and metropolitan area networks part 16: air interface for broadband wireless access systems amendment 1: multiple relay specification, 2009.

[3] J. Boyer, D.D. Falconer, H. Yanikomeroglu, Multihop diversity in wireless relaying channels, IEEE Transactions on Communications 52 (10) (2004) 1820–1830.

[4] J.N. Laneman, D.N.C. Tse, G.W. Wornell, Cooperative diversity in wireless networks: efficient protocols and outage behavior, IEEE Transactions on Information Theory 50 (12) (2004) 3062–3080.

[5] A.K. Sadek, W. Su, K.J.R. Liu, Multinode cooperative communications in wireless networks, IEEE Transactions on Signal Processing 55 (1) (2007) 341–355.

[6] V. Genc, S. Murphy, Y. Yang, J. Murphy, IEEE 802.16j relay-based wireless access networks: an overview, Proceedings of IEEE Wireless Communications 15 (5) (2008) 56–63.

[7] L. Erwu, W. Dongyao, L. Jimin, S. Gang, J. Shan, Performance evaluation of bandwidth allocation in 802.16j mobile multi-hop relay networks, Proceedings IEEE Vehicular Technology Conference (2007) 939–943.

[8] V. Genc, S. Murphy, J. Murphy, Performance analysis of transparent relays in 802.16j MMR networks, in: Proceedings of International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks, 2008, pp. 273–281.

[9] E. Visotsky, J. Bae, R. Peterson, R. Berry, M.L. Honig, On the uplink capacity of an 802.16j system, in: Proceedings of IEEE Wireless Communications and Networking Conference, 2008, pp. 2657–2662.

[10] H. Wang, C. Xiong, V.B. Iversen, Uplink capacity of multi-class IEEE 802.16j relay networks with adaptive modulation and coding, in: Proceedings of IEEE International Conference on Communications, 2009, pp. 1–6.

[11] B. Lin, P.H. Ho, L.L. Xie, X. Shen, Relay station placement in IEEE 802.16j dual-relay MMR networks, in: Proceedings of IEEE International Conference on Communications, 2008, pp. 3437–3441.

[12] H.C. Lu, W. Liao, Joint base station and relay station placement for IEEE 802.16j networks, in: Proceedings of IEEE Global Telecommunications Conference, 2009, pp. 1–5.

[13] D. Niyato, E. Hossain, D.I. Kim, Z. Han, Joint optimization of placement and bandwidth reservation for relays in IEEE 802.16j mobile multihop networks, in: Proceedings of IEEE International Conference on Communications, 2009, pp. 1–5.

[14] C.Y. Chang, C.T. Chang, M.H. Li, C.H. Chang, A novel relay placement mechanism for capacity enhancement in IEEE 802.16j WiMAX networks, in: Proceedings of IEEE International Conference on Communications, 2009, pp. 1–5.

[15] Y. Ge, S. Wen, Y.H. Ang, Analysis of optimal relay selection in IEEE 802.16 multihop relay networks, in: Proceedings of IEEE Wireless Communications and Networking Conference, 2009, pp. 1–6.

[16] W.H. Kuo, Recipient maximization routing scheme for multicast over IEEE 802.16j relay networks, in: Proceedings of IEEE International Conference on Communications, 2009, pp. 1–6.

[17] Y. Ge, S. Wen, Y.N. Ang, Y.C. Liang, Optimal relay selection in IEEE 802.16j multihop relay vehicular networks, IEEE Transactions on Vehicular Technology (2010) 2198–2206.

[18] K. Sundaresan, S. Rangarajan, Efficient algorithms for leveraging spatial reuse in OFDMA relay networks, Proceedings of IEEE INFOCOM (2009) 1539–1547.

[19] W. Wang, Z. Guo, J. Cai, X.S. Shen, C. Chen, Multiple frequency reuse schemes in the two-hop IEEE 802.16j wireless relay networks with asymmetrical topology, Computer Communications 32 (11) (2009) 1298–1305.

[20] D. Niyato, E. Hossain, D. Kim, Z. Han, Relay-centric radio resource management and network planning in IEEE 802.16j mobile multihop relay networks, IEEE Transactions on Wireless Communications 8 (12) (2009) 6115–6125.

[21] W. Wang, C. Chen, Z. Guo, J. Cai, X.S. Shen, Isolation band based frequency reuse scheme for IEEE 802.16j wireless relay networks, in: Proceedings of ICST International Conference on Heterogeneous Networking for Quality, Reliability, Security and Robustness, 2008, pp. 1–7.

[22] V. Genc, S. Murphy, J. Murphy, An interference-aware analytical model for performance analysis of transparent mode 802.16j systems, IEEE Broadband Wireless Access Workshop (BWA) co-located with GLOBECOM, 2008, pp. 1–6.

[23] V. Genc, S. Murphy, J. Murphy, Analysis of transparent mode IEEE 802.16j system performance with varying numbers of relays and associated transmit power, in: Proceedings of IEEE Wireless Communications and Networking Conference, 2009, pp. 1–6.

[24] J.P. Yoon, W.J. Kim, J.Y. Baek, Y.J. Suh, Efficient uplink resource allocation for power saving in IEEE 802.16 OFDMA systems, in: Proceedings of IEEE Vehicular Technology Conference, 2008, pp. 2167–2171.

[25] J. Snyman, Practical Mathematical Optimization: An Introduction to Basic Optimization Theory and Classical and New Gradient-based Algorithms, Springer-Verlag, 2005.

[26] B. Rong, Y. Qian, H.H. Chen, Adaptive power allocation and call admission control in multiservice WiMAX access networks, IEEE Wireless Communications 14 (1) (2007) 14–19.

[27] A.F. Molisch, Wireless Communications, Wiley, 2005.

[28] N. Bulusu, J. Heidemann, D. Estrin, GPS-less low-cost outdoor localization for very small devices, IEEE Personal Communications 7 (5) (2000) 28–34.

[29] M. Bshara, U. Orguner, F. Gustafsson, L. Van Biesen, Fingerprinting localization in wireless networks based on received-signal-strength measurements: a case study on wimax networks, IEEE Transactions on Vehicular Technology 59 (1) (2010) 283–294.

[30] H. Kellerer, U. Pferschy, D. Pisinger, Knapsack Problems, Springer-Verlag, 2004.

[31] C. Bettstetter, G. Resta, P. Santi, The node distribution of the random waypoint mobility model for wireless ad hoc networks, IEEE Transactions on Mobile Computing 2 (3) (2003) 257–269.