

RESEARCH

Open Access



User association for energy harvesting relay stations in cellular networks

Zhe Wang¹, Xiaodong Wang^{1*}, Motasem Aldiab² and Tareq Jaber²

Abstract

We consider a cellular wireless network enhanced by relay stations that are powered by renewable energy sources. Such a network consists of the macro base stations (BS), relay stations (RSs), and many mobile stations (MSs). In addition to the traditional data/voice transmission between the BS and the MSs, a higher service tier may be provided by using the energy harvesting RSs for some MSs. We propose a network scenario utilizing the energy harvesting relay stations to improve the service quality without taking the additional licensed frequency band and transmission power, and design a user association algorithm for the energy harvesting RSs in such a network. The goal is to assign each MS an RS for relaying its signal to minimize the probability of the relay service outage, i.e., the probability that an MS's relay service request is rejected. First, we propose a network scenario and develop a mathematical model to estimate the rejection probability for a given user association. We then propose a low-complexity local search algorithm, which balances the computational complexity and the performance, to obtain a locally optimal user association. Simulation results are provided to demonstrate the superior performance of the proposed techniques over the traditional methods.

Keywords: Cellular network, Energy harvesting, Relay station, Markov chain, User association, Queueing model

1 Introduction

Presently, broadband data services have become an increasingly significant source for mobile operators' businesses [1]. Such broadband services require much higher data throughput performance than the traditional voice services to meet the end users' quality-of-service (QoS) requirements, which bring challenges to the existing cellular wireless infrastructure [2]. One effective and low-cost solution is to employ relay stations in the existing cellular network. Moreover, with the rapid development of energy harvesting technologies, a new paradigm of wireless communications that employs energy harvesting transmitters has become a reality [3, 4]. In particular, energy harvesting relay stations are of great interest since they can improve service quality of the cellular network without taking the additional resources of the macro base station.

Typically, a relay station is a low-power and flexibly deployable base station that is able to provide complementary wireless access to the macrocell network within

a short range [5]. In both the WiMAX and the long-term evolution (LTE) systems, the deployment of relay station is a cost-effective solution to improving downlink data rate as well as to extending cell coverages [2, 6]. In particular, when the relay station communicates with the nearby mobile stations using a low transmission power and only uses the non-conflict unlicensed band, it can work as a network enhancement device to provide enhanced service without taking additional resources of the macro base station, e.g., the frequency band and the transmission power. On the other hand, green communication is a new evolution trend in the telecommunication industry, which calls for the use of renewable energy, and improved energy efficiency [7]. The rapid development of the renewable energy technology makes the deployment of the energy harvesting (EH) relay station possible [8]. Utilizing the renewable energy, e.g., the energy harvested from the sun or wind, the EH relay station will not only have the flexible deployment capability but also consumes less amount of the traditional energy, leading to lower carbon emission [5, 9]. To optimally utilize the renewable energy, scheduling for the energy harvesting transmitters is also discussed [10–12].

*Correspondence: wangx@ee.columbia.edu

¹Electrical Engineering Department, Columbia University, New York, NY 10027, USA

Full list of author information is available at the end of the article

Due to the significant difference to the traditional femto station powered by the grid and connected to the backbone network, the new EH relay station and the corresponding network operations need to be carefully designed by incorporating the energy budget due to the limited energy harvesting capability and the channel budget due to the limited capacity of the wireless macro base station-relay station link. In [13], the resource management was discussed based on a relay model from the fairness perspective; and in [14], distributed resource allocation algorithms were proposed for coexisting femto- and macrocell networks under an interference model. With the consideration of energy saving, the handover policy was studied for LTE networks with the use of femto stations [15]. However, the resource management problem under the renewable energy source becomes even more challenging under the QoS constraint, due to the dynamic nature of the energy harvesting process. In [16], the optimal energy usage was discussed for the base stations (BSs) powered by both on-grid energy and renewable energy. A cooperative system with the energy harvesting relays was analyzed in [17]. The authors of [18] studied the resource management problem for a mesh topology network with renewable energy source based on a queueing model, taking into account the relay path selection and the admission control. In [19], the energy management for a relay system was considered from the efficiency perspective under the battery constraints. In this paper, we consider a relay-assisted cellular network where the relay stations are powered solely by renewable energy sources and address the user association problem under the constraints of the energy and channel resources.

We assume that the relay station is small and light so that it can be deployed/redeployed flexibly, indoor or outdoor, according to the users' demands. Specifically, the voice service and the uplink data transmission are provided by the *macro base station* via the direct links between the BS and the *mobile stations* (MSs). For the downlink data transmission service, it can be provided either by the BS via the direct link or by the EH *relay station* (RS) via the relay link (i.e., BS-RS-MS). In particular, due to the channel condition, the best feasible service tier (or the best feasible transmission rate) provided by the relay link may be higher than that provided by the directly link for some MSs. In this case, when the demand tier of the MS cannot be fulfilled via the direct BS-MS link, the relay service is requested and then will be admitted whenever the RS's energy and channel resources are enough. We use the probability that a relay service request is rejected by the RS, i.e., *rejection probability*, to evaluate the performance of the relay service. In addition, any rejected relay service request would be redirected to the BS for the best available tier service. Moreover, we assume that the BS is powered by the grid and has enough

channel and energy resources but the energy and channel resources of the EH RS are limited due to the energy harvesting and wireless BS-RS link, respectively. Also, we assume that the RS uses the non-conflict unlicensed frequency band for the RS-MS transmission and may serve multiple MSs concurrently.

In case of RS coverage overlap, the user association can be introduced to optimize the quality of the RS service, i.e., to let each MS register at certain RS to maximize the overall network service quality. Such a user association problem can be formulated as a stochastic combinatorial optimization problem. Solving such a problem is extremely difficult in case that the profiles of the RSs, e.g., their resource limitations and the channel conditions, are not identical. We first develop a model to describe the relationship between the user association and the rejection probability, based on a queueing model and a Markov chain model. Using this model, the original stochastic optimization problem is converted to a deterministic combinatorial optimization problem. To solve the problem, a local search method is proposed to balance the performance and the complexity. Specifically, the user association specifies the RS assigned to each MS for relaying service in the subsequent time slots, which is static over certain period and updated by the proposed algorithm when a new MS enters/leaves the BS's coverage area or the profile of some MS changes. Simulation results are provided to demonstrate the effectiveness of the proposed method.

The remainder of the paper is organized as follows. In Section 2, we describe the system of the hierarchical relay-assisted cellular network and propose the network scenario. In Section 3, we provide a mathematical model that describes the relationship between the user association and the service quality. In Section 4, we provide a local search method to find the near-optimal user association. Simulation results are provided in Section 5. Finally, Section 6 concludes the paper.

2 System descriptions

2.1 Network scenario

We consider a conventional cellular network consisting of BSs and MSs, where the BS is connected to the backbone network via cable or fiber and powered by the grid, and the MS may be a cellphone, a tablet, or a laptop, etc., and is powered by battery. The BS and MS can communicate with each other in duplex mode on the licensed frequency band with the proper channel-access control. We assume that the BS has enough energy and bandwidth resources.

We now introduce energy harvesting RSs as the network enhancement devices, which are solely powered by some renewable energy source (along with a buffer battery) and equipped with the high gain antennas, relaying part of the downlink data transmission. We assume that

the transmissions from BS to RS, and from RS to MS, are simplex, where the BS-RS transmission uses the licensed frequency band, and the RS-BS transmission uses the non-conflict unlicensed frequency band, e.g., the available frequency band detected by cognitive radio techniques or the frequency band assigned to another cell. Since the high-gain antennas are equipped by the RS, a higher data transmission speed can be achieved for the transmission from BS to RS with the limited bandwidth and transmission power, as compared to the BS-MS link. Also, when the MS is close to the BS, the high speed transmission from the MS to the BS can also be achieved without taking any additional licensed frequency band by using a low transmission power. As a result, as compared with the BS-MS link, assisted by the EH RS, multiple MSs may be served simultaneously with the bandwidth and transmission power used for a single BS-MS link, achieving higher downlink transmission speed, i.e., improve the service quality without taking the additional resources of the BS.

We assume that each RS is associated with a fixed BS, i.e., relays the downlink transmission from the associated BS, and each MS is associated with a designated RS, i.e., receives the data from the associated RS. When a MS enters the BS's coverage area, it registers at the BS and searches the nearby RSs which is associated to the BS. For each discovered RS, if the channel condition of the RS-MS link is better than that of the BS-MS link by certain degree, we consider it as a feasible RS, i.e., by using the RS, a high service tier becomes available without using the additional

licensed frequency band. In particular, for the MS, multiple RSs may be available. In order to give a proper RS-MS association, the MS needs to submit its profiles, including the channel conditions between it and different feasible RSs and the expected arrival rate of the (downlink) relay service requests, to the BS.

Note that since the RS is solely powered by the renewable energy source, its incoming energy is limited, and we assume a simple structure and operation mode for the RS. Also, due to the use of the non-conflict unlicensed frequency band, the RS can only serve the MSs in a limited range and the channel conditions may be significantly better than those of the BS-MS links. Then, we can assume that the RS provides the relay service for all associated MSs at the same and fixed service tier (transmission rate), e.g., the top tier. Moreover, due to the bottleneck of the BS-RS link and the availability of the non-conflict unlicensed frequency band, the RS may only serve a limited number of the MSs concurrently, i.e., the RS has limited channel resource for RS-MS links. Then, we assume that the interference between different RS-MS links can be neglected.

The structure of the RS-assisted cellular network is given in Fig. 1.

2.2 User association

When a new MS enters/leaves the BS's coverage area or the profile of some MS changes, the BS generates a new *user association*, indicating the RS assigned to each MS for relaying service in the subsequent time slots. We denote

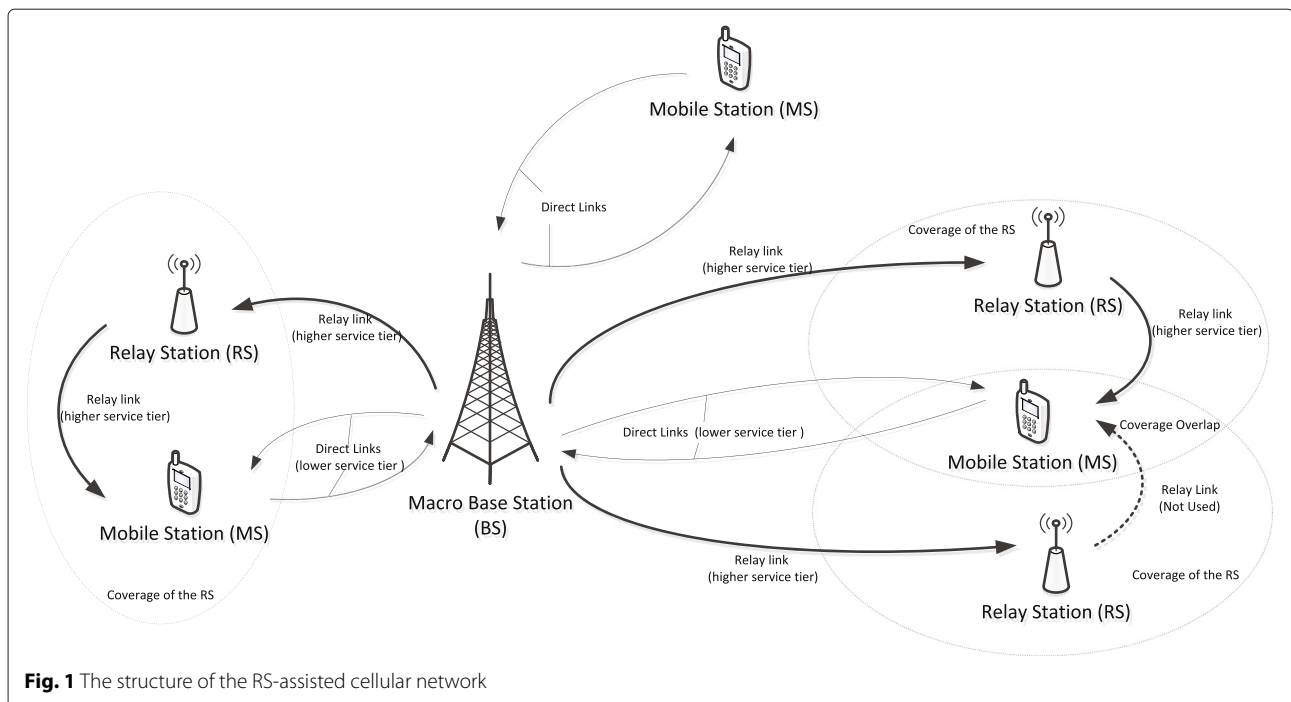


Fig. 1 The structure of the RS-assisted cellular network

$j \in \{1, 2, \dots, J\}$ as the index of each MS registered at the BS and $i \in \{1, 2, \dots, I\}$ as the index of each RS connected to the BS. We define $\mathbf{s} = [s_1, s_2, \dots, s_J]$ as the user association where s_j is the index of the assigned RS for MS j .

We assume that the channel conditions of different RS-MS pairs (i, j) may be different and denote $P_{i,j}$ as the discrete energy consumption level per time slot to achieve the required service tier. For each RS-MS pair, let P_i denote the individual maximum energy consumption per time slots by RS i . Then, if $P_{i,j} > P_i$, the required service tier cannot be achieved by the RS; therefore, we consider that RS i is not feasible for MS j and cannot be assigned to RS i , i.e., $s_j \neq i$. The set of feasible user association can then be defined as

$$\mathcal{S}(\mathcal{P}) \triangleq \{ \mathbf{s} : P_{s_j,j} < P_{s_j}, (P_{s_j,j}, P_{s_j}) \in \mathcal{P}, j = 1, 2, \dots, J \}, \tag{1}$$

where $\mathcal{P} \triangleq \{P_{i,j}, P_i \mid i = 1, 2, \dots, I, j = 1, 2, \dots, J\}$.

Note that the user association is static over a certain number of time slots and is updated when a new MS enters/leaves the BS's coverage area or some MS's profile changes. Once the user association is updated, the BS immediately informs the MSs and RSs to follow the new user association. Although the MS may be covered by multiple RSs, each MS can only be assigned to one

RS. Moreover, since the BS collects all of the information about the associated RSs and the MSs in its coverage area, including the parameters of the RS's channel and energy resources, the channel condition of each RS-MS pair, and the MS's transmission profile, we can perform the algorithms in the BS to obtain the user association and then send them to the related RSs and MSs, as shown in Fig. 2.

2.3 MS admission control

After an MS is assigned an RS, the relay service request of the MS is then handled by the assigned RS. In particular, if there are some large data packages queueing for transmitting to the MS from the BS and the direct link cannot provide the required service quality, the relay service is requested by the MS for these packages, and then the RS decides whether the request will be admitted or rejected, based on the channel and energy resource availabilities. If the request is admitted, these packages are transmitted from the BS to the MS via the assigned RS at high service tier. Otherwise, the packages have to be transmitted through the regular BS-MS link at the best available service tier, which is lower than that of the relay service.

To start an relay service, the MS j needs to initiate an *relay service request* ("request" for short in the remainder of the paper) first, indicating the requested transmission

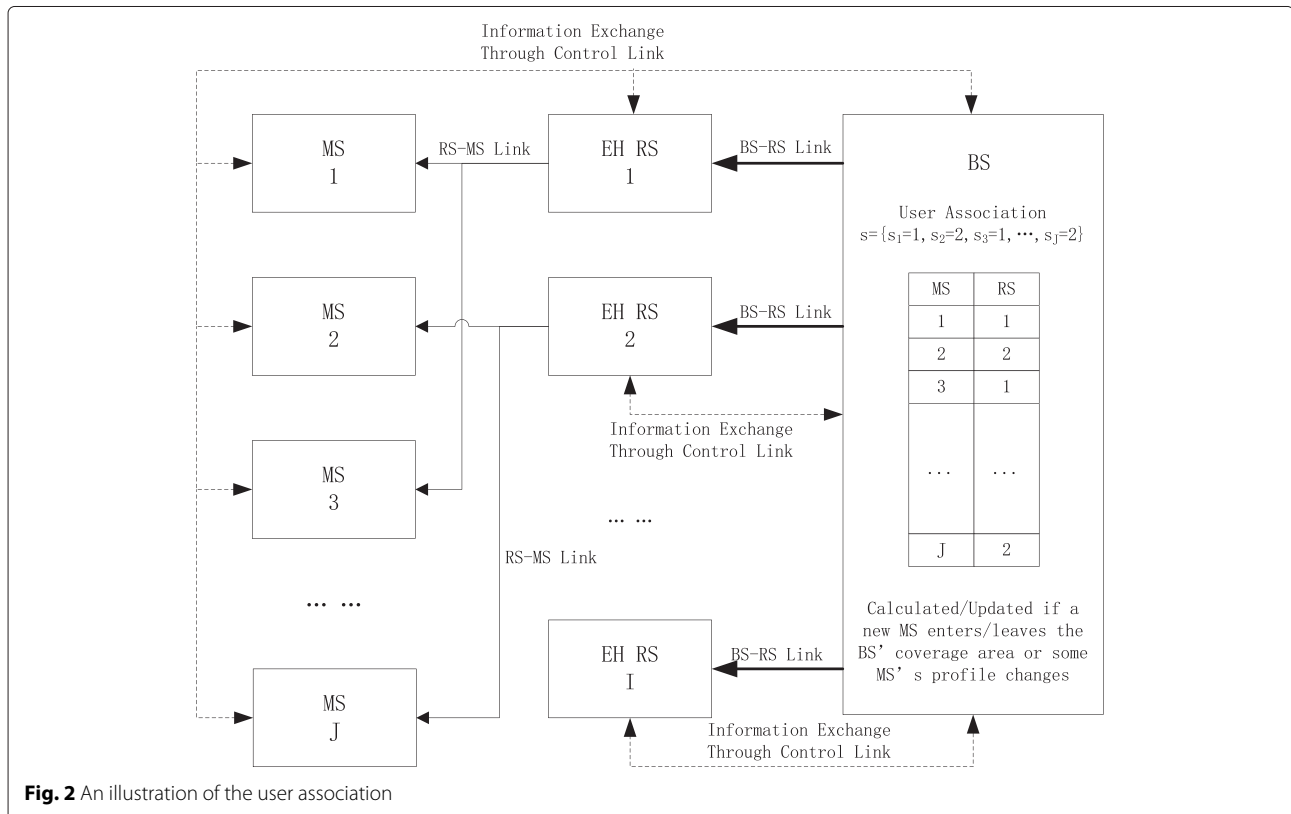


Fig. 2 An illustration of the user association

duration, denoted by D_j^k , where k is the index of the time slot when the request is made. In addition, we use $D_j^k = 0$ to indicate that no request is made in the k th time slot. The relay service request is processed by the assigned RS and an admission control decision is made instantaneously. Specifically, whenever both the uncommitted *channel resource* and *energy resource* of the RS are sufficient to fulfill the MS's relay service request, this request is admitted. Otherwise, the request is rejected.

The RS's channel resource, denoted by C_i , restricts the number of concurrently served MSs. Denote the channel resource state vector for RS i by $\mathbf{c}_i = [c_i^1, c_i^2, \dots, c_i^K]$ where c_i^k is the number of MSs that RS i is committed to serve at time slot k . Then, in order for RS i to admit the request $D_j^k > 0$, we must have

$$c_i^\ell < C_i, \ell = k+1, k+2, \dots, k + \lceil D_j^k/T \rceil, \quad (2)$$

where T is the duration of a time slot. If the request D_j^k is admitted, \mathbf{c}_i needs to be updated immediately as follows:

$$c_i^\ell \leftarrow c_i^\ell + 1, \ell = k+1, k+2, \dots, k + \lceil D_j^k/T \rceil. \quad (3)$$

The RS should also have sufficient energy to admit an MS relay service request. Assuming that the stochastic energy harvesting process is stationary within the assignment interval, e.g., K time slots, we denote E_i^k as the amount of the energy harvested by RS i in time slot k . Note that, since the energy harvesting is a stochastic process, to guarantee the service reliability for the admitted request, we evaluate the energy availability in terms of the *uncommitted energy* in each slot. Specifically, upon admitting a relay service request by the RS, certain amount of energy is committed, and we denote the uncommitted energy level stored in the battery of RS i at the beginning of the k th time slot by B_i^k , where $0 \leq B_i^k \leq B_i^{\max}$, and B_i^{\max} is the maximum allowed uncommitted energy, which is mainly determined by the battery capacity. Moreover, we denote B_i^0 as the initial energy in the battery.

Another necessary condition for admitting the request $D_j^k > 0$ by RS i is then

$$B_i^k \geq P_{ij} \lceil D_j^k/T \rceil. \quad (4)$$

If the request D_j^k is admitted, then we update the uncommitted energy as follows:

$$B_i^{k+1} = \min \left\{ B_i^k - P_{ij} \lceil D_j^k/T \rceil + E_i^k, B_i^{\max} \right\}. \quad (5)$$

Moreover, at the end of each time slot when no request is admitted, i.e., when $D_j^k > 0$ is rejected or $D_j^k = 0$, B_i^k is updated according to

$$B_i^{k+1} = \min \left\{ B_i^k + E_i^k, B_i^{\max} \right\}. \quad (6)$$

In summary, (2) and (4) constitute the necessary and sufficient conditions for RS i to admit the request $D_j^k > 0$ by MS j at slot k , where $s_j = i$.

3 System modeling

We assume that the energy harvesting E_i^k is a discrete independent and identically distributed (i.i.d.) random variable, whose probability density function (PDF) is denoted as $\text{Prob}(E_i^k = E)$, where $E \in \{0, 1, \dots, E_i^{\max}\}$ is the harvested energy and E_i^{\max} is the maximum energy harvesting capability of RS i . We further assume that the number of relay service requests made by MS j in a time slot follows a Poisson distribution with rate λ_j (requests per time slot), i.e.,

$$\text{Prob}(m \text{ requests by MS } j \text{ in a slot}) = \frac{\lambda_j^m}{m!} e^{-\lambda_j}. \quad (7)$$

Also, since the RSs provide the same service tier for all MSs, e.g., the top tier, we assume that the requested transmission duration D_j^k follows an exponential distribution with the average transmission duration μ , i.e., the PDF of D_j^k is given by

$$f_D(x) = \mu e^{-\mu x}. \quad (8)$$

Each relay service request $D_j^k > 0$ by MS j is processed by the pre-assigned RS s_j and is either admitted or rejected based on the channel and energy resource availabilities. Our objective is to design the user association \mathbf{s} to minimize the average probability $R(\mathbf{s})$ of a relay service request by an MS in the cellular network being rejected, where $R(\mathbf{s}) \triangleq \text{Prob}(D_j^k > 0 \text{ is rejected}, \forall j, k)$. In what follows, we will develop models that lead to an approximate expression for $R(\mathbf{s})$.

3.1 Model decomposition

Define \mathcal{R} as the event that a relay service request by an MS is rejected. Further define \mathcal{R}_c and \mathcal{R}_e as the events that a relay service request is rejected due to conditions (2) and (4) that are violated, respectively. Then, we have $\mathcal{R} = \mathcal{R}_c \cup \mathcal{R}_e$. By supposing that the channel resource or energy resource is unlimited first, we can get two independent models, the model with unlimited energy resource (UE model), where $\mathcal{R} = \mathcal{R}_c$, and the model with unlimited channel resource (UC model), where $\mathcal{R} = \mathcal{R}_e$.

First, we suppose that the energy resource is unlimited and we have the UE model. For a specific RS i , its probability of rejecting a relay service request can be obtained by resorting to an M/M/S/S queueing model [20, 21], where the customer arrival process follows the Poisson distribution, the service time follows the exponential distribution, there are S servers, and there is no waiting room.

Since the request arrivals of MS j follow the Poisson distribution with the arrival rate λ_j , the request arrivals at

each RS also follow a Poisson distribution with the arrival rate [20]

$$\tilde{\lambda}_i(\mathbf{s}, \alpha_c) = \alpha_c \sum_{j \in \mathcal{J}_i(\mathbf{s})} \lambda_j, \tag{9}$$

where α_c is a discount parameter ($\alpha_c = 1$ when the energy resource is unlimited) and $\mathcal{J}_i(\mathbf{s}) \triangleq \{j \mid s_j = i\}$ is the set of assigned MSs to RS i . Moreover, since the transmission duration follows the exponential distribution, so does the service time in this queueing model. Also, the channel resource capacity C_i characterizes the number of servers.

Then, according to this queueing model, the rejection probability caused by the violation of (2) in RS i corresponds to the probability of a new service request arrival when all servers are busy and is given by [20, 21],

$$R_c^i(\mathbf{s}, \alpha_c) = \left[\frac{\tilde{\lambda}_i(\mathbf{s}, \alpha_c)}{\mu} \right]^{C_i} / \left[C_i! \sum_{c=0}^{C_i} \frac{\tilde{\lambda}_i(\mathbf{s}, \alpha_c)^c}{\mu^c c!} \right]. \tag{10}$$

Next, suppose that the channel resource is unlimited and we have the UC model. We use a Markov chain to model the evolution of the uncommitted energy at RS i , as shown in Fig. 3. Denote $\mathcal{B}_i \triangleq \{0, 1, \dots, B_i^{\max}\}$ as the set of the states, where each state $m \in \mathcal{B}_i$ represents the uncommitted energy level in the battery. According to (5), when a request is admitted, the energy required to fulfill this request is committed. Then, after each time slot, the uncommitted energy level and therefore the state of the chain may change depending on whether a relay service request is admitted or not, based only on the state, i.e., the uncommitted energy, in the last time slot. In particular, if the energy consumption of the requested transmission is larger than the uncommitted energy in the

battery, the request would be rejected. We next compute the probability of this event.

Since the duration of the time slot T and the request arrival rate λ_i are small enough, the probability that more than one relay service requests are made in a time slot is small, e.g., when $T = 100$ ms, $1/\lambda = 100$ s/request, $\text{Prob}(\text{request} = 1)/\text{Prob}(\text{request} > 1) > 5 \times 10^4$, and $\text{Prob}(\text{request} > 1) < 5 \times 10^{-7}$. We note that, even if at most one relay service request is made in a time slot, the RS can serve multiple MSs currently since a service request corresponds to a transmission that may last several time slots. Then, in the UC model, we make the approximate assumption that at most one relay service request is made in a time slot. The probability that MS j makes a relay service request to RS s_j in a time slot is then

$$p_j(\mathbf{s}, \alpha_e) = 1 - \text{Prob}(0 \text{ request by MS } j \text{ in a slot}) = 1 - e^{-\alpha_e \lambda_j} \tag{11}$$

where α_e is a discount factor ($\alpha_e = 1$ when the channel resource is unlimited).

Since D_j^k follows the exponential distribution and the probability that the request duration D_j^k is d time slots is

$$\begin{aligned} \text{Prob}(\lceil D_j^k / T \rceil = d) &= \int_{(d-1)T}^{dT} \mu e^{-\mu x} dx \\ &= e^{-\mu dT} (e^{\mu T} - 1). \end{aligned} \tag{12}$$

Moreover, given specific average transmission duration μ , we can find d^{\max} such that $\text{Prob}(D_n^k > d^{\max} T)$ is small enough. Then, we assume d^{\max} as the maximum requested transmission duration in the UC model. Specifically, the MS can only submit the request with $D_i^k = d^{\max} T$ even if more time slots are required to complete the package transmission. Mapping the MS's request to the

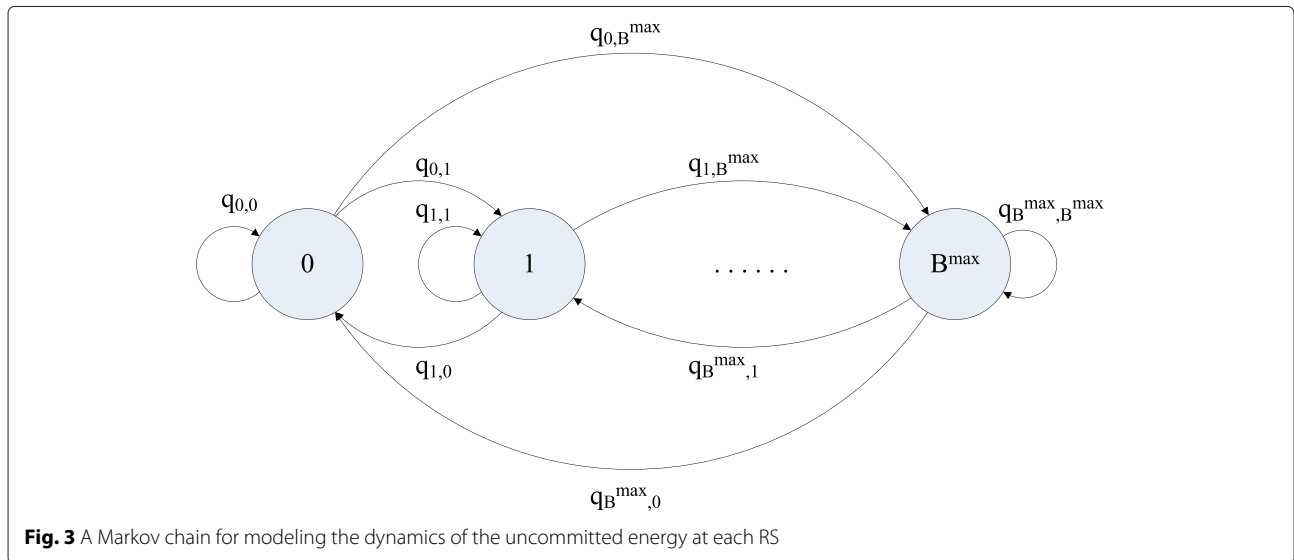


Fig. 3 A Markov chain for modeling the dynamics of the uncommitted energy at each RS

energy commitment, the probability that w units of energy are committed for MS j 's request in a time slot is given by

$$\beta_j(w) = \begin{cases} e^{-\mu dT} (e^\mu - 1), & \text{if } w = P_{s_j} d, d = 1, 2, \dots, d^{\max} - 1, \\ e^{-\mu dT}, & \text{if } w = P_{s_j} d, d = d^{\max}, \\ 0, & \text{otherwise.} \end{cases} \quad (13)$$

Then, we can further write the probability that w units of energy are committed by RS i in a time slot as

$$\gamma_i(\mathbf{s}, w, \alpha_e) = \sum_{j \in \mathcal{J}_i(\mathbf{s})} p_j(\mathbf{s}, \alpha_e) \beta_j(w). \quad (14)$$

Define the state transition probability matrix as $\mathbf{Q}_i(\mathbf{s}, \alpha_e) \triangleq [q_{m,n}^i(\mathbf{s}, \alpha_e)] \in [0, 1]^{(B_i^{\max}+1) \times (B_i^{\max}+1)}$, where $q_{m,n}^i(\mathbf{s}, \alpha_e)$ is the state transition probability for RS i . Since the state of the Markov chain represents the uncommitted energy level in the battery, its state transition follows the energy commitment process in (4)–(5), with the transition probabilities given by (14). In particular, at the end of each time slot, if the energy commitment w is such that the uncommitted energy exceeds the battery capacity, i.e., $B_i^k + E_i^k - w > B_i^{\max}$, the chain transits to state B_i^{\max} , corresponding to energy overflow. On the other hand, if w is such that $B_i^k + E_i^k - w < 0$, the chain transits to state $\min \{B_i^{\max}, B_i^k + E_i^k\}$, corresponding to request rejection.

Note that, for each RS, we can form a Markov chain, whose transition probability is dependent on two i.i.d. random variables, i.e., the committed energy w and the harvested energy E . Specifically, for given any pair of states m and n , $q_{m,n}^i(\mathbf{s}, \alpha_e)$ is the sum probabilities of all possible energy combination of w and E that leads the chain to transit from state m to state n . Then, the state transition probability matrix $\mathbf{Q}_i(\mathbf{s}, \alpha_e)$ can be calculated as follows.

Algorithm 1 Calculate the transition probability matrix $\mathbf{Q}_i(\mathbf{s}, \alpha_e)$

1: Initialization

$$\mathbf{Q}_i(\mathbf{s}, \alpha_e) = \mathbf{0}$$

2: Element Update

FOR $m = 0, 1, \dots, B_i^{\max}$

FOR all possible realizations of $(w, E) \in$

$$\{0, P_{s_j} 2 \cdot P_{s_j}, \dots, d^{\max} \cdot P_{s_j}\} \times \{0, 1, \dots, E_{s_j}^{\max}\}$$

IF $m \geq w$

$$n = \min \{B_i^{\max}, m - w + E\}$$

ELSE

$$n = \min \{B_i^{\max}, m + E\}$$

ENDIF

$$q_{m,n}^i(\mathbf{s}, \alpha_e) \leftarrow q_{m,n}^i(\mathbf{s}, \alpha_e) + \gamma_i(\mathbf{s}, w, \alpha_e) \cdot$$

$$\text{Prob}(E_i^k = E)$$

ENDFOR

ENDIF

In the above procedures, we initially set the state transition matrix as $\mathbf{0}$. For each starting state $m = 0, 1, 2, \dots, B_i^{\max}$, we calculate the ending state and the corresponding transition probability based on the MS admission control rules with all possible combinations of the realizations of the random variables. Specifically, for a specific starting state m and the realizations (w, E) . We first check if the current battery level can afford the committed energy w of the current relay service request. If so, the request is admitted and then the state is transit to n following (5); otherwise, the request is rejected and then the state is transit to n following (6). Also, the state transition is associated with a probability, which is the product of $\gamma_i(\mathbf{s}, w, \alpha_e)$ and $\text{Prob}(E_i^k = E)$. Finally, we accumulate the probabilities for all possible (w, E) that leads the chain to transit from state m to state n .

Moreover, for a special case that the energy harvesting process is static, i.e., $E_i = E$, the only random variable in the chain is w . Then, using the above method, we can easily give the elements of the state transition matrix in an explicit form:

$$q_{m,n}^i(\mathbf{s}, \alpha_e) = \begin{cases} \gamma_i(\mathbf{s}, m - n + E, \alpha_e), & m - n \geq -E, n \neq B_i^{\max}, n \neq 0, \\ \sum_{w=m+E}^{W_i} \gamma_i(\mathbf{s}, w, \alpha_e), & m - n \geq -E, n = 0, \\ \sum_{w=0}^{m+E-B_i^{\max}} \gamma_i(\mathbf{s}, w, \alpha_e), & m - n \geq -E, n = B_i^{\max}, \\ 0, & \text{otherwise,} \end{cases} \quad (15)$$

where $W_i \triangleq P_i d^{\max}$ is the maximum possible committed energy of RS i for each relay service request.

Since this finite-state Markov chain is irreducible and aperiodic, there exists a stationary distribution [22], denoted by $\boldsymbol{\pi}_i(\mathbf{s}, \alpha_e) = [\pi_0^i(\mathbf{s}, \alpha_e), \pi_1^i(\mathbf{s}, \alpha_e), \dots, \pi_{B_i^{\max}}^i(\mathbf{s}, \alpha_e)]$ that can be obtained by solving the equation $\boldsymbol{\pi}_i = \boldsymbol{\pi}_i \mathbf{Q}_i$. Then, given a user association \mathbf{s} , the probability that an MS relay service request is declined by RS i due to the shortage of uncommitted energy, is given by

$$R_e^i(\mathbf{s}, \alpha_e) = \frac{1}{\gamma_i(\mathbf{s}, 0, \alpha_e)} \sum_{m=0}^{B_i^{\max}} \pi_m^i(\mathbf{s}, \alpha_e) \sum_{w=m+1}^{w \leq W_i} \gamma_i(\mathbf{s}, w, \alpha_e), \quad (16)$$

where $\gamma_i(\mathbf{s}, 0, \alpha_e)$ is the probability that no relay service request is made in a time slot.

3.2 Rejection probability estimation

We find that the events \mathcal{R}_e and \mathcal{R}_c are highly correlated only in case that multiple energy-consuming requests are made within a short period, leading to the exhaustion of both channel and energy resources. However, if the RS has adequate resources, e.g., the channel capacity, the battery capacity and energy harvesting capability are high, we

may assume that \mathcal{R}_e and \mathcal{R}_c are approximately independent. Indeed, in order to maintain an acceptable quality of service and system reliability, the RSs need to be equipped with sufficient resources. Then, we have

$$R^i(\mathbf{s}) = R_c^i(\mathbf{s}, \alpha_c) + R_e^i(\mathbf{s}, \alpha_e) - R_c^i(\mathbf{s}, \alpha_c)R_e^i(\mathbf{s}, \alpha_e), \quad (17)$$

with the proper discount factors α_c and α_e .

When \mathcal{R}_e (or \mathcal{R}_c) occurs, a relay service request is rejected, which can be viewed as equivalent to removing a request from the relay service request arrival process, without committing the energy or channel resource. To approximate the model with \mathcal{R}_e and \mathcal{R}_c , we consider such removal, which is caused by \mathcal{R}_e (or \mathcal{R}_c), as a process that randomly sampling the relay service request arrival, resulting in another Poisson process with a discounted arrival rate [20], where the discount factor is denoted by α_c (or α_e). Specifically, since we assume that \mathcal{R}_e and \mathcal{R}_c are approximately independent, when the energy resource and the channel resource are both limited, i.e., a request may be rejected by \mathcal{R}_e , \mathcal{R}_c , or both, we may use the probability that a request is not solely rejected by \mathcal{R}_e (or \mathcal{R}_c) to characterize the discount factor α_c (or α_e), i.e.,

$$\alpha_c = 1 - R_e^i(\mathbf{s}, \alpha_e) + R_c^i(\mathbf{s}, \alpha_c)R_e^i(\mathbf{s}, \alpha_e), \quad (18)$$

and

$$\alpha_e = 1 - R_c^i(\mathbf{s}, \alpha_c) + R_e^i(\mathbf{s}, \alpha_e)R_c^i(\mathbf{s}, \alpha_c). \quad (19)$$

By solving the fixed-point Eqs. (18)–(19), we can obtain the values of $R_c^i(\mathbf{s}, \alpha_c)$ and $R_e^i(\mathbf{s}, \alpha_e)$. Finally, an estimate of the probability that a relay service request is rejected is given by

$$R(\mathbf{s}) = \sum_{i=1}^I \frac{\sum_{j \in \mathcal{J}_i(\mathbf{s})} \lambda_j}{\sum_{j=1}^J \lambda_j} R^i(\mathbf{s}). \quad (20)$$

Finally, we summarize the procedure for estimating the rejection probability as follows.

Algorithm 2 Rejection probability estimation given \mathbf{s}

1: **FOR** $i = 1, 2, \dots, I$

 Set $\alpha_e^{(0)} = 1$ and $\alpha_c^{(0)} = 1$;

 Set $m = 1$;

DO

 Calculate $R_c^i(\mathbf{s}, \alpha_c^{(m-1)})$ by (10);

 Calculate $R_e^i(\mathbf{s}, \alpha_e^{(m-1)})$ by (16);

 Update α_c^m, α_e^m by (18)–(19);

UNTIL $\alpha_c^{(m)}$ and $\alpha_e^{(m)}$ converge

 Calculate $R^i(\mathbf{s})$ by (17);

ENDFOR

2: Calculate $R(\mathbf{s})$ by (20);

4 Computing the user association

We note that the probability that a relay service request is rejected in (20) is a function of the user association \mathbf{s} . We would like to find the optimal user association that minimizes the rejection probability, i.e.,

$$\mathbf{s}^* = \arg \min_{\mathbf{s} \in \mathcal{S}(\mathcal{P})} R(\mathbf{s}). \quad (21)$$

Specifically, the user association is static over a certain number of time slots, and we solve the problem in (21) when the user association requires to be updated, e.g., a new MS enters/leaves the BS's coverage area or some MS's profile changes.

Since the problem in (21) is a deterministic combinatorial optimization problem, it can be solved using the exhaustive search over the set of feasible assignments $\mathcal{S}(\mathcal{P})$. To perform the exhaustive search, we can first generate the entire set of the feasible user associations $\mathcal{S}(\mathcal{P})$ in (1) for the given channel conditions \mathcal{P} . Then, we estimate the rejection probability $R(\mathbf{s})$ for each user association $\mathbf{s} \in \mathcal{S}(\mathcal{P})$. Finally, the user association \mathbf{s}^* with the minimum $R(\mathbf{s})$ is the optimal one.

4.1 Local search algorithm

When the cardinality of the feasible set $\mathcal{S}(\mathcal{P})$ is large, the complexity of the exhaustive search becomes prohibitive. To strike a balance between the computational complexity and the performance, we propose a low-complexity local search algorithm to compute the sub-optimal user association. The basic idea is to start from a feasible assignment and, in each iteration, consider assignments that are within a distance d to the current best assignment and pick the one with the lowest rejection probability, until no improvement can be made.

We define the distance between two assignments, \mathbf{s}^1 and \mathbf{s}^2 , denoted by $D(\mathbf{s}^1, \mathbf{s}^2)$, as the Hamming distance between \mathbf{s}^1 and \mathbf{s}^2 , i.e.,

$$D(\mathbf{s}^1, \mathbf{s}^2) \triangleq \sum_{j=1}^J \mathbb{I}(s_j^1 \neq s_j^2) \quad (22)$$

and define $\mathcal{F}(\mathbf{s}, d, \mathcal{P}) \triangleq \{\tilde{\mathbf{s}} \mid D(\tilde{\mathbf{s}}, \mathbf{s}) \leq d, \tilde{\mathbf{s}} \in \mathcal{S}(\mathcal{P})\}$ as the set of feasible user associations with distance to \mathbf{s} no more than d . Starting from an initial feasible user association $\mathbf{s}^{(0)}$, our proposed algorithm iteratively performs the following two steps: *assignment composition* that generates the local feasible set $\mathcal{F}(\mathbf{s}, d, \mathcal{P})$ based on the current feasible assignments and *assignment selection* that evaluates the rejection probabilities of the assignments in the local feasible set and selects the one with the lowest rejection probability, i.e.,

$$\tilde{\mathbf{s}}^* = \arg \min_{\tilde{\mathbf{s}}} \{R(\tilde{\mathbf{s}}) \mid \tilde{\mathbf{s}} \in \mathcal{F}(\mathbf{s}, d, \mathcal{P})\}, \quad (23)$$

until $R(\tilde{\mathbf{s}}^*)$ no longer changes.

Specifically, in the assignment composition step, a local feasible set $\mathcal{F}(\mathbf{s}, d, \mathcal{P})$ needs to be generated. Denote

$$\mathcal{U}(d) \triangleq \{U \mid U \subseteq \{1, 2, \dots, J\}, |U| \leq d\} \quad (24)$$

as the set of MS groups with no more than d MSs. For each $U \in \mathcal{U}(d)$, we can try all possible user association changes for the MSs in U . For each change, if the resulting user association is still in $\mathcal{S}(\mathcal{P})$, this assignment can be added to $\mathcal{F}(\mathbf{s}, d, \mathcal{P})$. In addition, since different local feasible sets may contain the common user associations, to avoid the repeated calculation, we can cache the result of each calculated user association in a lookup table, especially when d is large.

The local search algorithm is summarized as follows.

Algorithm 3 Local search algorithm

1: Initialization

Set a feasible user association $\mathbf{s}^{(0)} \in \mathcal{S}(\mathcal{P})$;

Set $m = 1$, specify the maximum number of iterations M ;

2: Assignment composition

Generate $\mathcal{F}(\mathbf{s}^{(m-1)}, d, \mathcal{P})$;

Calculate $R(\mathbf{s})$ for each $\mathbf{s} \in \mathcal{F}(\mathbf{s}^{(m-1)}, d, \mathcal{P})$ using the rejection probability estimation algorithm;

3: Assignment selection

Update $\mathbf{s}^{(m)} = \arg \min \{R(\tilde{\mathbf{s}}) \mid \tilde{\mathbf{s}} \in \mathcal{F}(\mathbf{s}, d, \mathcal{P})\}$;

IF $R(\mathbf{s}^{(m-1)}) - R(\mathbf{s}^{(m)}) > 0$ and $m \leq M$

$m \leftarrow m + 1$;

GOTO STEP 2;

ENDIF

Note that the distance parameter d in the above algorithm controls the balance between the system performance and computational complexity. For example, when $d = 1$, we search only among the assignments that differ from the current solution by one element. As d increases, the search space grows with improved performance. When $d = J$, the algorithm becomes the exhaustive search that finds the optimal assignment.

4.2 Computational complexity

Denote \mathcal{P} as the cardinality of the set \mathcal{P} , i.e., the number of elements in \mathcal{P} . We define $F_j \triangleq |\{P_{ij} \mid P_{ij} \leq P_i, i = 1, 2, \dots, J\}|$. We use the number of operations in calculating $R(\mathbf{s})$ to measure the computational complexity. In the local search algorithm, in each iteration, $(|\mathcal{F}(\mathbf{s}, d, \mathcal{P})| - 1)$ user associations are evaluated and the number of iterations is upper bounded by $M(|\mathcal{F}(\mathbf{s}, d, \mathcal{P})| - 1)$. On the other hand, since totally there are $\prod_{j=1}^J F_j$ different feasible user associations, the computational complexity,

i.e., the maximum number of $R(\mathbf{s})$ calculations is bounded by

$$T(d) \leq \min \left(\prod_{j=1}^J F_j, M(|\mathcal{F}(\mathbf{s}, d, \mathcal{P})| - 1) \right). \quad (25)$$

Moreover, given the distance parameter d and the initial user association \mathbf{s} , we can change the user association at most d different MSs for local search in each iteration. There are $\binom{J}{d}$ possible MS subsets U and each can form at most $\prod_{j \in U} F_j - 1$ different user associations. Denoting $F_{\max} \triangleq \max \{F_j \mid j \in \mathcal{J}\}$, we then have the following relaxations,

$$\prod_{j \in U} F_j - 1 \leq F_{\max}^d, \text{ for any } U, \quad (26)$$

and

$$\begin{aligned} |\mathcal{F}(\mathbf{s}, d, \mathcal{P})| &\leq \sum_{U \in \mathcal{U}(d)} \left(\prod_{j \in U} F_j - 1 \right) \\ &\leq \sum_{U \in \mathcal{U}(d)} F_{\max}^d \\ &\leq \binom{J}{d} F_{\max}^d. \end{aligned} \quad (27)$$

Then, (25) can be further upper bounded by

$$\begin{aligned} T(d) &\leq \min \left(\prod_{j=1}^J F_j, M \binom{J}{d} F_{\max}^d \right) \\ &\leq \min \left(\prod_{j=1}^J F_j, MJ! F_{\max}^d \right) \\ &\sim \mathcal{O} \left(F_{\max}^d \right). \end{aligned} \quad (28)$$

Specifically, when we set $d = 1$, we have $T(1) \leq \mathcal{O}(F_{\max})$, and we obtain a sub-optimal assignment. On the other hand, when we set $d = J$, we have $T(J) \leq \mathcal{O}(F_{\max}^J)$, which is the complexity of the exhaustive search. By adjusting the parameter d , we can balance the system performance and the computational complexity.

5 Simulation results

We consider a sub-network of the cellular system consisting of $I = 3$ RSs and $J = 7$ MSs with the request arrival rates $\lambda_1 = 0.009, \lambda_2 = 0.008, \lambda_3 = 0.006, \lambda_4 = 0.005, \lambda_5 = 0.004, \lambda_6 = 0.003$, and $\lambda_7 = 0.002$ requests per time slot, the average transmission duration parameter $\mu = 0.05$ requests per time slot, and the maximum requested duration is $D^{\max} = 250$ time slots. For each RS, we assume that the battery capacity is $B_i^{\max} = 200$ units, the initial energy level is $B_i^0 = 100$ units, the energy harvesting capacity E_i^k follows the truncated Gaussian distribution (between 0 and 20) with the mean of 10 and

the variance of 2, and the power threshold is $P_i = 4$ units per time slot. For the channel set realizations $\mathcal{P} = \{P_i^j, P_i \mid i = 1, 2, \dots, I, j = 1, 2, \dots, J\}$, we generate them following the uniform distribution between 1 and 4 and pick up the ones with mean between [2.45, 2.55] for simulations. In addition, we set $d = 1$ in the proposed local search algorithm and the simulation horizon is set as 10^7 time slots.

For performance comparison purpose, we consider two simple user association strategies: the *load-balanced* strategy where the MSs are assigned to each RS (nearly) evenly regardless of the channel conditions, e.g., $\mathbf{s} = [1, 1, 1, 2, 2, 3, 3]$ when there are $J = 7$ MSs; the *best-channel* strategy where each MS is assigned to the RS with the best channel condition. Moreover, to evaluate the accuracy of the modeling techniques developed in Section 3, we also consider the “optimal assignment” which is obtained by exhaustively search for the user association that has the lowest rejection probability, where the rejection probability is evaluated through simulations rather than based on the proposed rejection probability estimation procedure.

To evaluate the performance, we randomly generate 100 sets of channel realizations and sort the simulation results in ascending order by the optimal rejection probability. In Fig. 4, the rejection probabilities of the assignments by the four methods are plotted. It is seen that, in most cases, the rejection probability of the assignment obtained by the proposed algorithm ($d = 1$) is the same or close to the optimal value while the rejection probabilities of the other two simple strategies are far away from the optimal one.

Moreover, we compare the performances of the optimal assignment, the model-based exhaustive search, the local search algorithm ($d = 1$), the load-balanced strategy, and the best-channel strategy, over different number of MSs, for $C_i = 2$ and $C_i = 3$, shown in Figs. 5 and 6, respectively. In this comparison, the performances are averaged over 200 sets of random channel realizations (we simulate 10^7 time slots for each channel realizations set). It is seen that the performance of the model-based exhaustive search almost coincides with that of the optimal assignment, while the tiny gap is caused by the approximation error of the rejection probability approximation model. We also see that the performance of the local search algorithm is only slightly worse than that of the exhaustive search. Compared to the load-balance assignment and the best channel assignment, a significant improvement is gained by using our proposed user association algorithm. Moreover, as compared in the two figures, the performances shown in Fig. 6 for the larger channel capacity $C_i = 3$ is better than that in Fig. 5 for the smaller channel capacity $C_i = 2$.

Recall that we use the number of the operations involved in the rejection probability estimation algorithm as the metric to measure the search complexity. Next, we compare the search complexity between our proposed algorithm and the exhaustive search. In this simulation, we consider that the first $J = 2, 3, 4, 5, 6, 7$ MS candidates are active, and we average the complexity, which is measured by the number of operations in calculating $R(\mathbf{s})$, over 200 random channel realizations. It is seen from Fig. 7 that our proposed local search algorithm has a significantly lower complexity than the exhaustive search method, especially when the number of MSs is large.

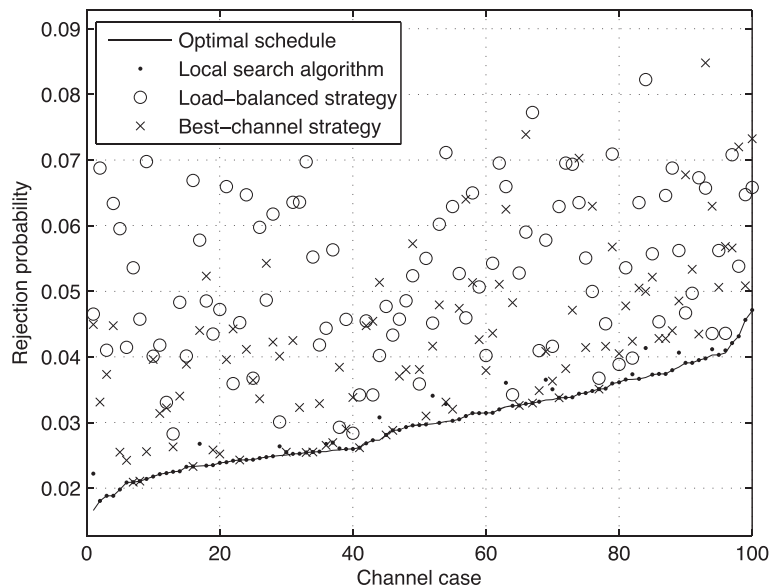


Fig. 4 Performance comparisons among different user associations ($J = 5, C_i = 2$)

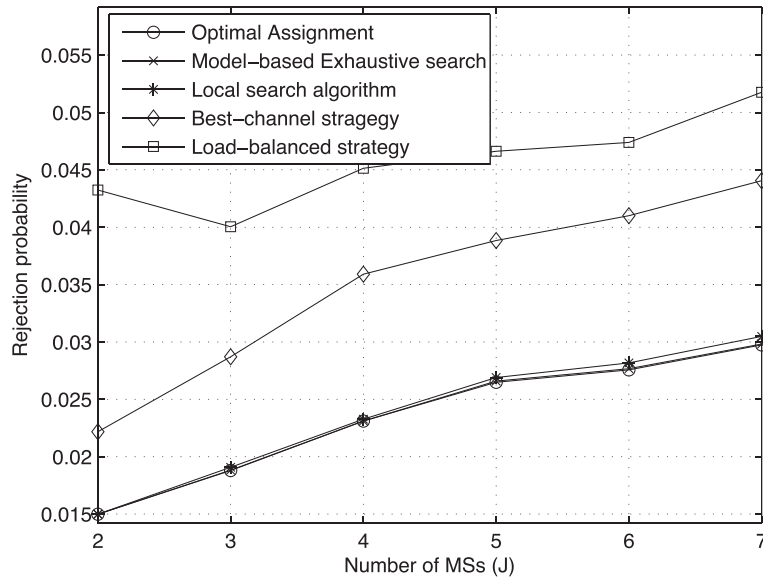


Fig. 5 Performance comparisons among different user associations over different number of MSs ($C_i = 2$)

6 Conclusions

We have considered a EH relay station to improve the service quality of the cellular network without taking the additional licensed frequency band and transmission power of the macro base station, where the RSs powered by renewable energy sources relay the downlink data transmissions from the macro base station to the mobile stations in some conditions. We have formulated the user association problem, which is to assign each MS an RS to

minimize the rejection probability of the requested relay service, as a combinatorial optimization problem based on a rejection probability estimation model. We proposed a local search algorithm to efficiently obtain a locally optimal assignment. Simulations have demonstrated that, even with the simplest local search ($d = 1$), the algorithm can still provide substantial service quality and reliability improvement over conventional methods. Finally, we note that the proposed framework can also be applied to an

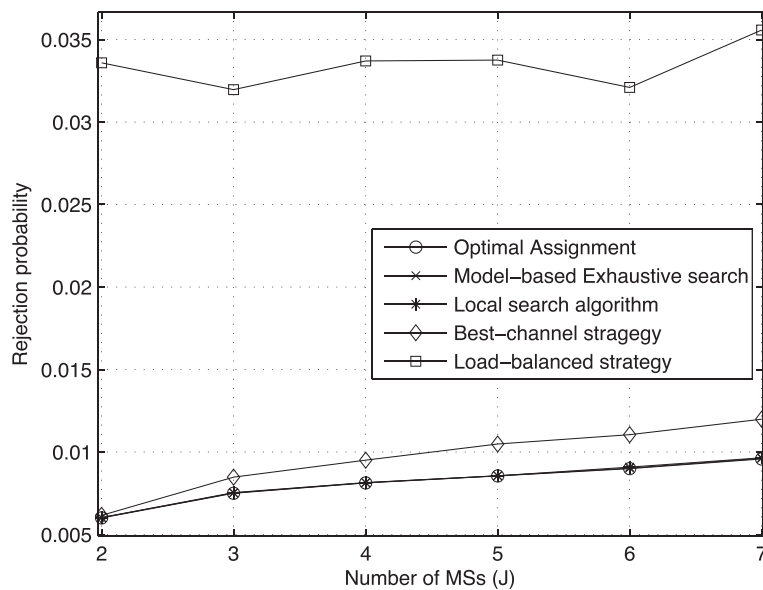


Fig. 6 Performance comparisons among different user associations over different number of MSs ($C_i = 3$)

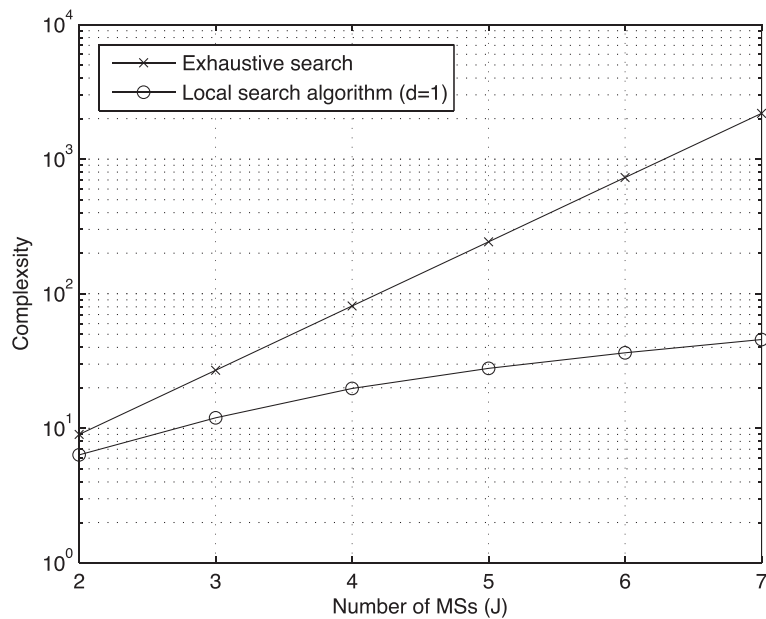


Fig. 7 Computational complexity comparison between the local search algorithm ($d = 1$) and the exhaustive search ($C_i = 2$)

heterogeneous network (HetNet) consisting of macrocells and picocells where the picocells are powered by energy harvesting devices.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Electrical Engineering Department, Columbia University, New York, NY 10027, USA. ²Faculty of Computing & Information Technology, University of Jeddah, Jeddah 21589, Saudi Arabia.

Received: 5 January 2015 Accepted: 1 December 2015

Published online: 23 December 2015

References

1. J Weitzen, T Grosch, Comparing coverage quality for femtocell and macrocell broadband data services. *IEEE Commun. Mag.* **48**(1), 40–44 (2010)
2. X Chu, Y Wu, D Lopez-Perez, X Tao, On providing downlink services in collocated spectrum-sharing macro and femto networks. *IEEE Trans. Wireless Commun.* **10**(12), 4306–4315 (2011)
3. JA Paradiso, T Starner, Energy scavenging for mobile and wireless electronics. *IEEE Trans. Pervasive Comput.* **4**, 18–27 (2005)
4. S Sudevalayam, P Kulkarni, Energy harvesting sensor nodes: survey and implications. *IEEE Commun. Surveys Tuts.* **13**(3), 443–461 (2011)
5. D Calin, H Claussen, H Uzunaloglu, On femto deployment architectures and macrocell offloading benefits in joint macro-femto deployments. *IEEE Commun. Mag.* **48**(1), 26–32 (2010)
6. Y Tseng, C Huang, Analysis of femto base station network deployment. *IEEE Trans. Veh. Technol.* **61**(2), 748–757 (2012)
7. S Yeh, in *Proc. 2010 Global Mobile Congress*. Green 4G communications: renewable-energy-based architectures and protocols (IEEE, Shanghai, P.R. China, 2010), pp. 1–5
8. Z Hasan, H Boostanimehr, V Hargava, Green cellular networks: a survey, some research issues and challenges. *IEEE Commun. Surveys Tuts.* **13**(4), 524–540 (2011)
9. A Fehske, G Fettweis, J Malmodin, G Biczok, The global footprint of mobile communications: the ecological and economic perspective. *IEEE Commun. Mag.* **49**(8), 55–62 (2011)
10. O Ozel, K Tutuncuoglu, J Yang, S Ulukus, A Yener, Transmission with Energy Harvesting Nodes in Fading Wireless Channels: Optimal Policies. *IEEE J. Sel. Areas Commun.* **29**(8), 1732–1743 (2011)
11. J Yang, S Ulukus, Optimal packet scheduling in an energy harvesting communication system. *IEEE Trans. Commun.* **60**(1), 220–230 (2011)
12. K Tutuncuoglu, A Yener, Optimum transmission policies for battery limited energy harvesting nodes. *IEEE Trans. Wireless Commun.* **21**(4), 1117–1130 (2013)
13. E Liu, Q Zhang, K Leung, Relay-assisted transmission with fairness constraint for cellular networks. *IEEE Trans. Mobile Comput.* **11**(2), 230–239 (2012)
14. Y Shi, A MacKenzie, in *Proc. GLOBECOM 2011*. Distributed algorithms for resource allocation in cellular networks with coexisting femto- and macrocells (IEEE, Houston, TX, USA, 2011), pp. 1–6
15. D Xenakis, N Passas, C Verikoukis, in *Proc. ICC 2012*. A novel handover decision policy for reducing power transmissions in the two-tier LTE network (IEEE, Ottawa, ON, Canada, 2012), pp. 1352–1356
16. T Han, N Ansari, On optimizing green energy utilization for cellular networks with hybrid energy supplies. *IEEE Trans. Wireless Commun.* **12**(8), 3872–3882 (2013)
17. B Medepally, N Mehta, Voluntary energy harvesting relays and selection in cooperative wireless networks. *IEEE Trans. Wireless Commun.* **9**(11), 3543–3553 (2010)
18. L Cai, Y Liu, T Luan, X Shen, J Mark, H Poor, in *Proc. GLOBECOM 2011*. Adaptive resource management in sustainable energy powered wireless mesh networks (IEEE, Houston, TX, USA, 2011), pp. 1–5
19. W Zhang, D Duan, L Yang, Relay selection from a battery energy efficiency perspective. *IEEE Trans. Commun.* **59**(6), 1525–1529 (2011)
20. S Ross, *Introduction to Probability Models*. (Academic Press, San Diego, CA, 2000)
21. G Giambene, *Queueing Theory and Telecommunications Networks and Applications*. (Springer, New York, 2005)
22. M Puterman, *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. (John Wiley & Sons, New York, 1994)