applied sciences

MDPI

Article

Power Control and Link Selection for Wireless Relay Networks with Hybrid Energy Sources

Runze Wu, Huan Xie *, Zhiyi Chen and Liangrui Tang

School of Electrical and Electronic Engineering, North China Electric Power University, Beijing 102206, China
* Correspondence: 1172201384@ncepu.edu.cn; Tel.: +86-132-6332-7756

Received: 9 June 2019; Accepted: 3 July 2019; Published: 6 July 2019

Abstract: The Hybrid energy supply (HES) wireless relay system is a new green network technology, where the source node is powered by the grid and relay is powered by harvested renewable energy. However, the network's performance may degrade due to the intermittent nature of renewable energy. In this paper, our purpose is to minimize grid energy consumption and maximize throughput. However, improving the throughput requires increasing the transmission power of the source node, which will lead to a higher grid energy consumption. Linear weighted summation method is used to turn the two conflicting objectives into a single objective. Link assignment and a power control strategy are adopted to maximize the total reward of the network. The problem is formulated as a discrete Markov decision model. In addition, a backwards induction method based on state deletion is proposed to reduce the computational complexity. Simulation results show that the proposed algorithm can effectively alleviate performance degradation caused by the lack of renewable energy, and present the trade-off between energy consumption and throughput.

Keywords: energy harvesting relay; energy consumption; throughput

1. Introduction

In order to expand the coverage of wireless networks and improve the communication quality of edge users, relay has been widely used in wireless networks such as long term evolution (LTE) and 5G [1,2]. However, dense deployment of relays can lead to problems such as high energy consumption and greenhouse gas emissions [3,4]. Energy harvesting technology is the most promising technology to solve the economic and environmental problems caused by the dense deployment of relays, which can collect and utilize renewable energy such as solar and wind energy [5–9]. In addition, energy harvesting technology can reduce the dependence of wireless networks on grid energy. Deploying green relays in areas where grid energy is scarce can effectively expand the coverage of wireless communication networks. However, the intermittent and random nature of the renewable energy may cause a decline in network performance. Therefore, it is critical to establish a renewable energy allocation and link selection mechanism to ensure the performance of wireless relay networks.

Optimal power control strategies to alleviate network performance degradation caused by the lack of renewable energy have been proposed in [10–13]. In [10], both the source and relay nodes are powered by renewable energy, and an off-line power control strategy is made to minimize data transmission time with throughput constraints. It was also proved that the power control strategy has water injection structure. Differing from [10], literature [11] considered how the source node is powered by the grid while the relay node is powered by renewable energy. An off-line power control strategy is proposed to maximize the grid energy efficiency of the source node. In [12], the research was extended to a Gauss fading channel, with both off-line and on-line power control schemes are proposed. In addition to the traditional relay, the literature [13] considers the relay with cache function. Off-line and on-line power control strategies are developed to maximize network throughput. All of

the above studies have developed off-line or on-line power control strategies under the constraints of user's quality of service (QoS).

Besides power control schemes, link selection strategies are also critical. In [14], the source node is powered by grid energy while the relay is powered by renewable energy. With causal side information, a link selection strategy is made to maximize the average throughput of all time slots. In [15], the relay is powered by radio frequency (RF) energy with a finite battery capacity, and only one data packet needs to be transmitted in each slot. A link selection scheme based on battery level is proposed to minimize the outage probability of the relay. Further, literature [16] jointly optimizes power and link selection to reduce outage probability of the relay. In [17], the research was extended to multi-relay with data caching, Under the constraints of battery capacity, on-line and off-line power allocation and link selection mechanisms were developed to minimize data transmission time.

Most previous studies on hybrid energy supply (HES) wireless relay systems assume that the amount of data transmitted by the source node are fixed in each slot, and then optimize the outage probability or transmission time. However, in practical applications, the source node needs to serve a lot of different users and deliver unfixed data bits in each slot. So, data bits should be transmitted as much as possible to reduce the network congestion with grid energy consumption constraints. Consequently, in this paper, we consider that the source node has available bits to be transmitted all the time. Our goal is to maximize network throughput while minimizing grid energy consumption.

The rest of this paper is organized as follows. Section 2 describes the system model. In Section 3, the Markov decision process (MDP) problem is formulated, for which a low-complexity algorithm is proposed. Simulations are shown in Section 4. Finally, Section 5 highlights the conclusion.

2. System Model

2.1. Network Model

We consider a green wireless relay network as shown in Figure 1, which consists of a source node (S), a relay node (R), and a destination node (D). The source node is powered by the electric grid, and its maximum transmit power is denoted as $p_S^{\max}$. The relay is powered purely by the renewable energy which is harvested from the nature and stored in a battery, with the maximum transmit power as $p_R^{\max}$. The distances between source node and relay, relay and destination node, source node and destination node are d_{SR}, d_{RD} and d_{SD}, respectively. We consider the performance changes of the network in N slots with length T. The slot index set is denoted as $\mathbb{N} = \{1, 2 \cdots\cdots, N\}$.

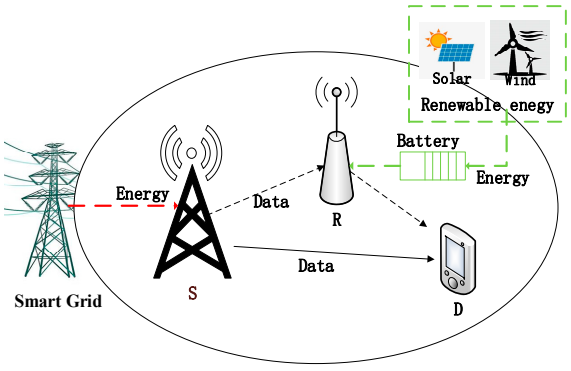

Figure 1. A green wireless network with an energy harvesting relay.

We assume that the source node has available bits to be transmitted all the time. The relay operates in a half-duplex manner that it receives data from the base station (BS) in the first half of the time slot and forwards it in the second half [11]. In each slot, there are two links for S to transmit bits: the relay link and direct link. The link assignment indicator in the i-th slot is denoted as $I_j^i \in \{0, 1\}$, with

$j \in \{SRD, SD\}$. $\mathrm{I}_{SRD}^{i} = 1$ indicates that the relay link is assigned to deliver the bits, and $\mathrm{I}_{SD}^{i} = 1$ represents the event that the direct link is selected. In each slot, only one of these links can be selected, so

$$\mathrm{I}_{SD}^{i} + \mathrm{I}_{SRD}^{i} = 1, \forall i \in \mathbb{N}. \tag{1}$$

In addition, transmission power of source node S and relay R in slot i is p_S^i and p_R^i, respectively. Static power consumption is neglected in this work.

We consider that the channel is time-varying with small-scale fading; denoted h_i^k, with $k \in \{SR, RD, SD\}$ as the channel fading factor between two nodes. In addition, all channels share the bandwidth (B) together. Therefore, once the direct link is selected in the i-th slot, the number of bits transmitted by the source node are calculated as

$$R_{SD}^{i} = BT \log(1 + h_{SD}^{i} g_0 d_{SD}^{-\alpha} \sigma^{-2} P_{SD}^{i}), \tag{2}$$

according to the Shannon theorem, where σ^{-2} is the noise variance of node D, g_0 is the channel fading constant, p_{SD}^i is the transmit power of node S and α is the path loss exponent. While the relay link is selected, the total bits delivered by the source node and the relay are given as

$$R_{SR}^{i} = B\frac{T}{2} \log(1 + h_{SR}^{i} g_0 d_{SR}^{-\alpha} \sigma^{-2} p_{SR}^{i}) \tag{3}$$

and

$$R_{RD}^{i} = B\frac{T}{2} \log(1 + h_{RD}^{i} g_0 d_{RD}^{-\alpha} \sigma^{-2} p_{R}^{i}), \tag{4}$$

where p_{SR}^i is the transmit power of node S when selecting the relay link.

2.2. Energy Model

In order to simplify the energy harvesting model, the harvesting process is thought to be accomplished at the beginning of each time slot [18]. Then, discrete energy model is adopted to describe the process of energy harvesting with E_H^i energy packets arriving at each time slot [19], which obeys the Poisson distribution with mean λ [20]. And λ represents the intensity of energy harvesting. In addition, each packet contains the energy of E_e. In the initial time slot, the energy stored in the battery is the original energy plus the harvested energy. While in the slot of $i > 1$, the energy consumed by the relay should be subtracted. Therefore, battery level in the i-th slot is expressed as:

$$E(i) = \begin{cases} E_0 + E_H^i E_e & i = 1 \\ E(i) - C_R^{i-1} + E_H^i E_e & i > 1 \end{cases}, \tag{5}$$

where E_0 is the initial energy, and C_R^{i-1} is energy consumed by the relay in the last slot, which is given by:

$$\mathrm{C}_R^{i-1} = \mathrm{I}_{SD}^{i-1} T \mathrm{p}_{SD}^{i-1} + \mathrm{I}_{SRD}^{i-1} \frac{T}{2} \mathrm{p}_{SRD}^{i-1}. \tag{6}$$

Since the battery is of the limited size, the following energy constraint should be satisfied:

$$E(i) \le E_{\max}, \tag{7}$$

where $E_{\max}$ is the maximum capacity of the battery. The energy consumed by the relay should be no more than the energy of the battery, that is $C_R^i \le E(i)$.

2.3. Optimizing Objective

Most previous studies on two-hop green wireless relay networks concentrated on minimizing the outage probability or transmission time with fixed bits at each slot [10,21,22]. However, when

the network is busy and needs to provide services for multiple users, there will be a continuous stream of data bits to be transmitted. In this case, throughput can show the carrying capacity of the network, which is attractive to us. In addition, the source node powered by the grid energy can also be applied to some wireless networks. For example, macro base stations powered by the grid energy can maintain basic coverage in heterogeneous wireless networks. To take above aspects into consideration, we set grid energy consumption and throughput as optimization objectives. However, according to Equations (2) and (3), improving the throughput requires increasing the transmission power of the source node, which will lead to higher grid energy consumption. Therefore, increasing throughput and reducing grid energy consumption are conflicting objectives, which can be formulated as:

$$obj = \begin{cases} \max \sum_{i=1}^{N} I_{SD}^{i} R_{SD}^{i} + I_{SRD}^{i} R_{RD}^{i} \\ \min \sum_{i=1}^{N} I_{SD}^{i} T p_{SD}^{i} + I_{SRD}^{i} \frac{T}{2} p_{SR}^{i} \end{cases}. \tag{8}$$

There are many ways to solve multi-objective problems, and linear weighted summation is an effective method of them [23]. We use the weighted summation method to transform the conflicting multi-objectives into a single objective. Thus, the conflict objectives can be denoted as:

$$TR = \sum_{i=1}^{N} [\omega_t (I_{SD}^{i} R_{SD}^{i} + I_{SRD}^{i} R_{RD}^{i}) - \omega_g (I_{SD}^{i} T p_{SD}^{i} + I_{SRD}^{i} \frac{T}{2} p_{SR}^{i})], \tag{9}$$

where ω_t and ω_g are weight coefficients of throughput and grid energy consumption, respectively. $\omega_t (I_{SD}^{i} R_{SD}^{i} + I_{SRD}^{i} R_{RD}^{i})$ and $-\omega_g (I_{SD}^{i} T p_{SD}^{i} + I_{SRD}^{i} \frac{T}{2} p_{SR}^{i})$ are the weighted throughput reward and grid energy reward in the i-th slot.

3. Optimal Control for Expected Total Rewards

In this section, we assume h_{SR}^{i}, h_{SD}^{i} and h_{RD}^{i} are causally known. Consequently, we aim to adapt the transmission power and link selection to maximize the expected total rewards. Thus, the problem can be formulated as:

$$\mathrm{P}_1 : \max_{I_j^i, p_{SR}^i, p_{SD}^i, p_R^i} E\left\{ \sum_{i=1}^{N} [\omega_t (I_{SD}^{i} R_{SD}^{i} + I_{SRD}^{i} R_{RD}^{i}) - \omega_g (I_{SD}^{i} T p_{SD}^{i} + I_{SRD}^{i} \frac{T}{2} p_{SR}^{i})] \right\} \tag{10}$$

$$\text{s.t. Equations (1) and (7)} \tag{11}$$

$$C_R^i \le E(i) \tag{12}$$

$$R_{SR}^{i} = R_{RD}^{i} \tag{13}$$

$$p_j^i \le p_j^{\max} \forall i \in \mathbb{N}, j \in \{S, R\} \tag{14}$$

$$I_j^i \in \{0, 1\}, \forall i \in N, j \in \{SD, SRD\}. \tag{15}$$

In P_1, the objective is the expected total rewards over N time slots. Equation (12) is the energy constraint that the energy consumed by relay should not exceed that of the battery. Equation (13) is the throughput constraint that the bits delivered in the two stages of relay link should be Equations (14) and (15) are the power constraint and the link selection constraint.

3.1. Problem Simplification

From Equation (10), we can see that the optimized variables of P_1 include the 0–1 variable I_j^i and the continuous variable p_{SR}^{i}, p_{SD}^{i} and p_R^i. Thus, it is very difficult to optimize so many different types

of variables at the same time. Therefore, we do some simplification of P_1 to reduce the optimized variables. And, the problem of P_1 can be expressed as:

$$P_2 : \max_{p_R^i} E\left\{\sum_{i=1}^{N}[\beta_i D_i - (1-\beta_i)H_i]\right\} \tag{16}$$

$$\text{s.t. } C_R^i \le E(i) \tag{17}$$

$$p_R^i \le p_R^{\max} \forall i \in \mathbb{N}. \tag{18}$$

And β_i is the 0–1 variable, which can be calculated as:

$$\beta_i = \begin{cases} 1 \; p_R^i \neq 0 \\ 0 \; p_R^i = 0 \end{cases} \quad \forall i \in \mathbb{N}. \tag{19}$$

That is, when the transmission power of relay is non-zero, $\beta_i = 1$ and relay link is chosen to transmit data. While the power of relay is zero, $\beta_i = 0$ and direct link is selected to deliver bits. In addition, D_i is the network total reward when relay link is chosen, which is given by:

$$\mathrm{D}_i = \omega_t R_{RD}^i - \omega_g G_{SR}^i, \tag{20}$$

where R_{RD}^i is the throughput, and G_{SR}^i is the grid energy consumed by the source node. We assume that the amount of data bits transmitted in the two stages of relay link should be equal. So, there is

$$R_{SR}^i = R_{RD}^i = B\frac{T}{2}\log(1 + h_{SR}^i g_0 d_{SR}^{-\alpha}\sigma^{-2}p_{SR}^i) = B\frac{T}{2}\log(1 + h_{RD}^i g_0 d_{RD}^{-\alpha}\sigma^{-2}p_R^i). \tag{21}$$

From equation (21), we can know that

$$h_{SR}^i d_{SR}^{-\alpha} p_{SR}^i = h_{RD}^i d_{RD}^{-\alpha} p_R^i. \tag{22}$$

According to Equation (20) to (22), D_i can be further given as:

$$D_i = w_t B\frac{T}{2}\log(1 + h_{RD}^i g_0 d_{RD}^{-\alpha}\sigma^{-2}p_R^i) - w_g\frac{T}{2}\frac{h_{RD}^i d_{RD}^{-\alpha} p_R^i}{h_{SR}^i d_{SR}^{-\alpha}}. \tag{23}$$

H_i is the reward when direct link is selected, and the problem of solving H_i can be expressed as:

$$P_H : \max_{p_{SD}^i} H_i = \max_{p_{SD}^i}[\omega_t R_{SD}^i - \omega_g G_{SD}^i] = \omega_t TB\log(1 + h_{SD}^i g_0' d_{SD}^{-\alpha}\sigma^{-2}p_{SD}^i) - \omega_g T p_{SD}^i \tag{24}$$

$$\text{s.t. } p_{SD}^i \le p_S^{\max} \forall i \in \mathbb{N}. \tag{25}$$

Proposition 1. *The reward of the direct link H_i is convex in transmission power p_S^i.*

Proof: The second derivative of formula (24) is

$$f(p_{SD}^i)'' = -\frac{w_t TB(h_{SD}^i g_0 d_{SD}^{-\alpha}\sigma^{-2})^2}{(1 + h_{SD}^i g_0 d_{SD}^{-\alpha}\sigma^{-2}p_{SD}^i)^2 \ln 2} < 0, \tag{26}$$

which means that $f(p_{SD}^i)$ is a convex function. □

According to the properties of convex functions, we can easily know that H_i gets the maximum value at $p_{SD}^i = \frac{\omega_t}{\omega_g} - \frac{1}{h_{SD}^i g_0 d_{SD}^{-\alpha} \sigma^{-2}}$. Therefore, the optimal p_{SD}^i and H_i are known once h_{SD}^i is given. However, p_{SD}^i has practical significance only on $[0, p_S^{\max}]$ in this work. Consequently, when $\frac{\omega_t}{\omega_g} - \frac{1}{h_{SD}^i g_0 d_{SD}^{-\alpha} \sigma^{-2}} < 0$, P_H decreases monotonously on $[0, p_S^{\max}]$ with maximum value $p_H(0)$. While $\frac{\omega_t}{\omega_g} - \frac{1}{h_{SD}^i g_0 d_{SD}^{-\alpha} \sigma^{-2}} > p_S^{\max}$, P_H increases monotonously on $[0, p_S^{\max}]$ with maximum value $p_H(\max)$.

It can be seen from Equation (16) to (26), that the values of β_i and D_i are determined by p_R^i, and the maximum value of H_i^* is a fixed value in each time slot which is only related to h_{SD}^i. Therefore, the optimal variable of problem P_2 is only p_R^i.

3.2. MDP Model for Expected Total Rewards

Our goal is to maximize the total rewards over N slots through a relay power control scheme. However, due to the limited battery capacity, the relay power selection results in each slot will affect the initial battery capacity at the next moment. So, power decisions on different time slots are mutually influential. The MDP is a useful model to handle such decision problems, and backward induction is an effective algorithm to solve this problem [24].

Therefore, we formulate P_2 as a Markov decision process (MDP) problem, which can also be expressed as:

$$\mathrm{P}_2 : \max_{\pi \in \prod} E^{\pi} \sum_{i=1}^{N} R(i), \tag{27}$$

where π is a feasible relay power policy, $\prod$ denotes the set of all feasible policies. $R(i)$ is the reward of slot i, which is given by $[\beta_i D_i - (1-\beta_i) H_i]$.

3.2.1. MDP Basics

A sequential decision-making method is the selection of one of several action strategies in each time slot during the operation of the system [25]. In the sequential decision process, if the transfer of the system state obeys the known probability law and is independent of the previous history, then this sequential problem is called an MDP problem [26]. An MDP model consists of a reward function, system states, actions, state transition probability and objective, each of which will be described in detail later.

3.2.2. Reward Function

In an MDP model, the reward function is defined as $r(j, a_i)$. It indicates that the system gets the reward with action a_i at state j [27]. This is denoted $o_i \in O, i \in \mathbb{N}$ as the rule for selecting relay power in slot i. Thus, the rules over N slots can be expressed as $\pi = (\mathrm{o}_1, \mathrm{o}_2, \mathrm{o}_3, \cdots \mathrm{o}_N)$, and the set of all possible rules is denoted as $\prod$. Given the initial state k and strategy $\pi \in \prod_m^O$, the expected total rewards can be also written as:

$$V_N(\pi, s_1) = \sum_{i=1}^{N} \sum_{a_i \in A_i, j \in S} P_\pi\{s_i = j, \Delta_i = a_i \big| s_1 = k\} r(j, a_i), \tag{28}$$

where s_i and Δ_i are the states of the relay system and the selected action in the slot i, respectively. $P_\pi\{s_i = j, \Delta_i = a_i \big| s_1 = k\}$ is the conditional probability of using strategy $\pi \in \prod_m^O$, starting from state k, selecting action a_i, and moving to state j at slot *i*. Our aim is to find the optimal action selection scheme as $\pi^* = \left(\mathrm{o}_1^*, \mathrm{o}_2^*, \cdots, o_N^*\right)$, which makes $V_N(\pi^*, s_1)$ the maximum value.

3.2.3. Discretization of System States and Actions

The optional values of the system states and actions should be finite in MDP model. However, the system states include links and the battery states, and the relay power actions are continuous values

in the wireless relay network. Therefore, it is necessary to discretize the system states and relay power actions. The relay system states consist of channel fading values and battery levels, which can be given as $s_i \triangleq \left\langle h^i_{SR}, h^i_{SD}, h^i_{RD}, \varepsilon(i) \right\rangle$. We discretize the channel states by reference to the method in literature [28]. Denoting $H = \{H_1, H_2, \ldots H_3\}$ as the set of channel fading values, which is an equal-difference sequence. The probability when the channel fading is H_k with $k \in \{1, 2 \cdots K\}$ can be calculated as:

$$p\left\{h^i_j = H_k\right\} = \frac{1}{K} \forall i \in \mathbb{N}, j \in \{SR, SD, SRD\}, k \in \{1, 2 \cdots K\}. \tag{29}$$

We divide the battery into M + 1 energy level. And the battery states set is taken as $\varepsilon(i) \in \varepsilon = [0, 1, \cdots, m, \cdots M]$. The real-time energy level of the battery can be calculated by

$$\varepsilon(i) = m = \left\lfloor \frac{E_i M}{E_{\max}} \right\rfloor. \tag{30}$$

Denote $A'_i = [0, \frac{p^{\max}_R}{L}, \ldots, p^{\max}_R]$ as the action set, which is also an equal-difference sequence. Actually, the value of the relay transmit power is constrained by the battery level. Thus, the action set is given as $A_i = [0, \frac{p^{\max}_R}{L}, \ldots, \min(p^{\max}_R, p^i_x)]$, where p^i_x is calculated by:

$$p^i_x = \left\lfloor \frac{2\varepsilon(i) E_{\max} L}{MT p^{\max}_R} \right\rfloor * \frac{p^{\max}_R}{L}. \tag{31}$$

$\lfloor x \rfloor$ is the function that rounds the variable x down.

3.2.4. State Transition Probability

After action a_i is selected, the system states will migrate from s_i to s_{i+1}, which can be expressed as $s_i \triangleq \left\langle h^i_{SR}, h^i_{SD}, h^i_{RD}, \varepsilon(i) \right\rangle \rightarrow s_{i+1} \triangleq \left\langle h^{i+1}_{SR}, h^{i+1}_{SD}, h^{i+1}_{RD}, \varepsilon(i+1) \right\rangle$. Since the value of channel fading is equal probability, the state transition probability is:

$$p\{s_{i+1} \big| s_i, a_i\} = \frac{1}{K^3} p\{\varepsilon_{i+1} \big| \varepsilon_i, a_i\}. \tag{32}$$

We assume that $\varepsilon(i)$ and $\varepsilon(i+1)$ are in the M_1 and M_2 levels of the battery, which should satisfy the Equation: $\frac{M_2 E_{\max}}{M} \leq \frac{M_1 E_{\max}}{M} - a_i * \frac{T}{2} + E^i_H E_e < \frac{(M_2+1) E_{\max}}{M}$. As mentioned in Section 2.2, the energy harvesting process obeys the Poisson distribution with mean λ. Therefore, Equation (32) can be further given as:

$$p\{s_{i+1} \big| s_i, a_i\} = \frac{1}{K^3} \sum_{n=n_1}^{n_2} \frac{\lambda^n}{n!} e^{-\lambda}, \tag{33}$$

where n_1 and n_2 is given as $n_1 = \left\lceil \frac{\frac{M_2 E_{\max}}{M} + a_i * \frac{T}{2} - \frac{M_1 E_{\max}}{M}}{E_e} \right\rceil$ and $n_2 = \left\lceil \frac{\frac{(M_2+1) E_{\max}}{M} + a_i - \frac{M_1 E_{\max}}{M}}{E_e} \right\rceil - 1$, respectively. $\lceil x \rceil$ is the function that rounds the variable x up.

3.3. The Backward Induction Algorithm for MDP Problem

The backward induction algorithm is an effective solution to the optimal strategy and value function in the finite-stage Markov decision programming problem [26]. A new function, $V^n_*(i)$, was proposed based on the backward induction algorithm, which is formulated as:

$$\begin{aligned} V^n_*(i) &= \max_{a_i \in A_i} [r(k, a_i) + \sum_{j \in s} p(j \big| k, a_i) V^{i+1}_*(j)] \\ &= r(k, f^i_*(k)) + \sum_{j \in s} p(j \big| k, a_i) V^{i+1}_*(j) \quad , \\ &(k \in s, i = \{N, N-1, N-2, \cdots, 0\}) \end{aligned} \tag{34}$$

where $V_*^{N+1}(k)=0, \forall k \in s$ [26]. According to Equation (34), the optimal value function of the expected total rewards can be calculated as $V_*^1 = (V_*^1(1), V_*^1(2), \cdots, V_*^1(q))$. Meanwhile, the decision sequence $\pi^* = (o_*^1, o_*^2, \cdots, o_*^N)$ obtained is the optimal strategy.

With the backward induction algorithm, the number of states required to traverse is $M \times L^3$. The state space may be very large if some of the elements are of large size and may encounter the curse of dimensionality [29]. An effective method to reduce the computational complexity in MDP model is proposed in literature [28]. In this case, we also eliminate some states that do not need to be searched according to the wireless relay network properties in our model by reference [28].

Proposition 2. *When h_{SR}^i, h_{RD}^i, $\varepsilon(i)$ are fixed value, and the optimal action is $a_i = 0$ at state $s_i^- \triangleq \left\langle h_{SR}^i, h_{SD}^{i^-}, h_{RD}^i, \varepsilon(i) \right\rangle$, the optimal action is $a_i = 0$ for any state s_i^+ of $h_{SD}^{i^+} > h_{SD}^{i^-}$.*

Proof: If the optimal action for the state s_i^- is $a_i = 0$, according to Equation (34), we know that:

$$r(s_i^-, 0) + \sum_{j \in s} p(j|s_i^-, 0)\, V_{i+1}^*(j) > \max_{a_i \in A_i and a_i \neq 0} \left[r(s_i^-, a_i) + \sum_{j \in s} p(j|s_i^-, a_i)\, V_{i+1}^*(j) \right]. \tag{35}$$

From Equation (21), we know that

$$r(s_i, 0) = H_i^* = \omega_t TB \log(1 + h_{SD}^i g_0' d_{SD}^{-\alpha} \sigma^{-2} {p_{SD}^i}^*) - \omega_g T {p_{SD}^i}^*, \tag{36}$$

Which becomes larger as h_{SD}^i grows. Therefore, for any state s_i^+ with $h_i^{SD^+} > h_i^{SD^-}$, $r(s_i^+, 0) > r(s_i^-, 0)$. Since h_i^{SR}, h_i^{RD} and ε_i are fixed value, we can get

$$\sum_{j \in s} p(j|s_i^+, 0)\, V_{i+1}^*(j) = \sum_{j \in s} p(j|s_i^-, 0)\, V_{i+1}^*(j), \tag{37}$$

and

$$\max_{a_i \in A_i and a_i \neq 0} \left[r(s_i^+, a_i) + \sum_{j \in s} p(j|s_i^+, a_i)\, V_{i+1}^*(j) \right] = \max_{a_i \in A_i and a_i \neq 0} \left[r(s_i^-, a_i) + \sum_{j \in s} p(j|s_i^-, a_i)\, V_{i+1}^*(j) \right]. \tag{38}$$

Finally,

$$r(s_i^+, 0) + \sum_{j \in s} p(j|s_i^+, 0) V_{i+1}^*(j) > \max_{a_i \in A_i and a_i \neq 0} \left[r(s_i^+, a_i) + \sum_{j \in s} p(j|s_i^+, a_i)\, V_{i+1}^*(j) \right], \tag{39}$$

which proves that the optimal action is $a_i = 0$ in the state s_i^+. □

Proposition 3. *When h_{SD}^i, h_{RD}^i, $\varepsilon(i)$ are fixed value, and the optimal action is $a_i = 0$ at state $s_i^+ \triangleq \left\langle h_{SR}^{i^+}, h_{SD}^i, h_{RD}^i, \varepsilon(i) \right\rangle$, the optimal action is $a_i = 0$ for any state s_i^- of $h_{SR}^{i^-} < h_{SR}^{i^+}$.*

Proof: If the optimal action for state s_i^+ is $a_i = 0$, according to Equation (34), we can get:

$$r(s_i^+, 0) + \sum_{j \in s} p(j|s_i^+, 0)\, V_{i+1}^*(j) > \max_{a_i \in A_i and a_i \neq 0} \left[r(s_i^+, a_i) + \sum_{j \in s} p(j|s_i^+, a_i)\, V_{i+1}^*(j) \right]. \tag{40}$$

While $p_R^i \neq 0$, $r(s_i, p_R^i)$ is given as

$$r(s_i, p_R^i) = D_i = w_t B \frac{T}{2} \log(1 + h_{RD}^i g_0 d_{RD}^{-\alpha} \sigma^{-2} p_R^i) - w_g \frac{T}{2} \frac{h_{RD}^i d_{RD}^{-\alpha} p_R^i}{h_{SR}^i d_{SR}^{-\alpha}}, \tag{41}$$

which becomes smaller as h_{SR}^{i} grows. Thus, For the stat s_i^- with $h_i^{SR^-} < h_i^{SR^+}$, $r(S_i^-, a_i) < (S_i^+, a_i) a_i \in A_i$ *and* $a_i \neq 0$. Since h_i^{SD}, h_i^{RD} and ε_i are fixed value, there are:

$$\max_{a_i \in A_i and a_i \neq 0} \left[r(s_i^-, a_i) + \sum_{j \in s} p(j \middle| s_i^-, a_i) V_{i+1}^*(j) \right] < \max_{a_i \in A_i and a_i \neq 0} \left[r(s_i^+, a_i) + \sum_{j \in s} p(j \middle| s_i^+, a_i) V_{i+1}^*(j) \right], \quad (42)$$

and

$$r(s_i^-, 0) + \sum_{j \in s} p(j \middle| s_i^-, 0) V_{i+1}^*(j) = r(s_i^+, 0) + \sum_{j \in s} p(j \middle| s_i^+, 0) V_{i+1}^*(j). \quad (43)$$

Finally,

$$r(s_i^-, 0) + \sum_{j \in s} p(j \middle| s_i^-, 0) V_{i+1}^*(j) = r(s_i^+, 0) + \sum_{j \in s} p(j \middle| s_i^+, 0) V_{i+1}^*(j), \quad (44)$$

which indicates that the optimal action is $a_i = 0$ in the state s_i^-.

Algorithm 1. Backward Induction Algorithm Based on States Elimination

Input: $p_R^{\max}$, $p_S^{\max}$, d_{SD}, d_{RD}, d_{SR}, T, B, N, K, L, λ, E_e, ω_t, ω_g, ε
Output: π^*
1: **Initialize** $\pi^* = zeros(N, K^3 \times M)$, $V_*^{N+1} = 0$
2: **While** N ≠ 1
3: $\mathrm{p}_x^N = \left\lfloor \frac{2\varepsilon(N)L}{T p_R^{\max}} \right\rfloor * \frac{p_R^{\max}}{L}$, $\mathrm{A}_N = [0, \frac{p_R^{\max}}{L}, \ldots, \min(p_R^{\max}, p_x^N)]$;
4: **For** m = 1 to M, k_{SR} = 1 to K
5: k_{RD} = K, k_{SD} = 1;
6: **While** $k_{SD} \neq K+1$, $k_{RD} \neq 0$
7: $s =< H_{k_{SR}}, H_{k_{SD}}, H_{k_{RD}}, \varepsilon(m) >$;
8: **For** j = 1: length (A_N)
9: **Calculate** $\pi^*(N, s) = \underset{A_N(j)}{\arg\max} V^N(s) = \underset{A_N(j)}{\arg\max} (r(s), A_N(j)) + \sum_{l \in s} p(l \middle| s, A_N(j)) V_*^{N+1}(l))$;
10: **End For**
11: **If** $\pi^*\left(N, s_{k_{SR}, k_{SD}, k_{RD}, m}\right) = 0$
12: $\pi^*(N, s_{k_{SR}, k_{SD}^+, k_{RD}, m}) = 0 \ \forall k_{SD}^+ > k_{SD}$, $\pi^*(N, s_{k_{SR}, k_{SD}, k_{RD}^-, m}) = 0 \ \forall k_{RD}^- < k_{RD}$;
13: $k_{RD} = k_{RD} - 1$, $k_{SD} = k_{SD} + 1$;
14: **Else**
15: **if** k_{SD} = K
16: $k_{RD} = k_{RD} - 1$, k_{SD} = 1;
17: **Else**
18: $k_{SD} = k_{SD} + 1$;
19: **End If**
20: **End If**
21: **End For**
23: **N = N − 1;**
24: **End While**

4. Numerical Simulations

In this section, we run some numerical simulations to analyze the total reward, grid energy consumption and throughput in two-hop wireless relay networks. In the simulations, we set B = 10 MHz, T = 1 ms, $p_S^{\max}$ = 2 W, $p_R^{\max}$ = 0.5 W, σ^2 = −97.5 dBm, g_0 = −40 dB, α = 4 [28]. And E_e = 0.01 mJ, $E_{\max}$ = 1.6 mJ, K = 10, L = 20, d_{SD} = 80 m, $\omega_g = 1$. The detailed numerical results are shown as follows.

4.1. Baseline Schemes

Joint Power Control and Link Selection Algorithm (JPLA): The JPCALSA only considers the current system state, and calculates the maximum rewards of relay and direct link, respectively. Then, the optimal access link is selected by comparing the rewards.

Power Control Algorithm (PCA): The PCA is to maximize the reward in single slot by adjusting the power of the relay, and link selection scheme is not taken into account [11].

When the energy of relay is sufficient, the system will be in the ideal state. In order to compare the ideal results with our results in different situations, we propose JPLA-F and BIABoSE-F, which are JPLA and BIABoSE with enough renewable energy.

4.2. Parameter Analysis

Figure 2 demonstrates the total rewards with different number of battery levels at different time slots. In any slot, the total rewards increase with the number of battery levels rises. Actually, the energy between two adjacent levels is expressed by the lower level, and the interval of two adjacent levels is smaller as the number of levels becomes larger. At this point, the error between the true value and the expressed value will be smaller, which makes a more accurate result. When the number of battery levels reach 80 and 160, their rewards are close and maximal. Consequently, for reducing the computational complexity, M = 80 is used for simulation analysis in the follow-up.

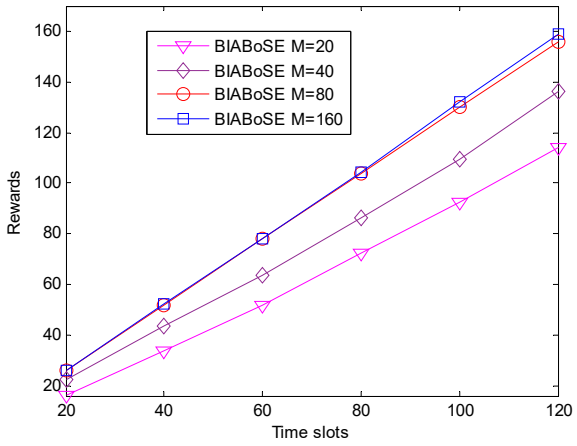

Figure 2. Total reward vs. time slots for different number of the battery intervals, $\omega_t = 1$, $d_{SD} = d_{SR} + d_{RD} = 40m$.

We assume that $d_{SD} = d_{SR} + d_{RD} = 80m$, and the total rewards vary with d_{SD} is shown in Figures 3 and 4. As d_{SD} increases, the rewards of all algorithms become larger first and then decrease. When d_{RD} is small, $d_{SR} = 80 - d_{RD}$ is large, and the path loss between the source node and relay is high. In this case, the source node delivers a few bits to relay with high grid energy consumption, which leads to low total rewards. As d_{RD} increases, the path loss between the source node and relay decreases, and the total rewards rise. Once d_{RD} is larger than a certain threshold, the path loss between the relay and destination is high, the number of bits that can be transmitted by relay is lower than that by source node. In this case, the total reward is gradually reduced as the throughput of relay tapers off.

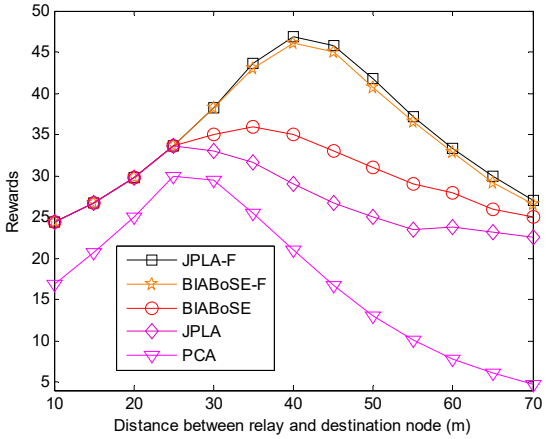

Figure 3. The total rewards versus d_{SD}, $\omega_t = 1$, $\lambda = 2$.

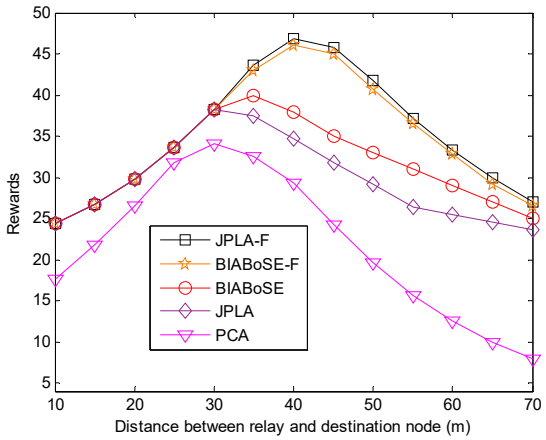

Figure 4. The total rewards versus d_{SD}, $\omega_t = 1$, $\lambda = 4$.

In addition, the JPLA-F and BIBAoSE-F achieve the maximum value near $d^*_{RD} = 40$ m in both Figures. Meanwhile, the PCA, JPLA and BIBAoSE obtain the maximum value at different d^*_{RD} in two Figures. Unlike the JPLA-F and BIBAoSE-F, the other three algorithms are affected by the energy harvesting intensity. The energy that the relay needed for data transmission grows larger as d_{RD} increases. Therefore, when the energy is more sufficient, the total rewards will be closer to optimal result. The total rewards reach the maximum value at $d_{RD} = 40$. Thus, we choose $d_{RD} = 40$ for subsequent simulations to better observe the improvement of system performance in the absence of energy.

4.3. Total Reward Maximization

Figure 5 shows the total rewards changes with the time slots. Compared with the PCA, the JPLA adds a link selection mechanism. Therefore, the JPLA can transmit data through the direct link when the battery is very low, which can increase the total rewards. The BIBAoSE takes the future system states into account, which makes a more efficient green energy allocation over N slots than the JPLA. However, all the algorithms can only alleviate the system performance degradation caused by insufficient energy and cannot replace the green energy supply. Therefore, the JPLA-F and the BIBAoSE-F always have the highest total rewards.

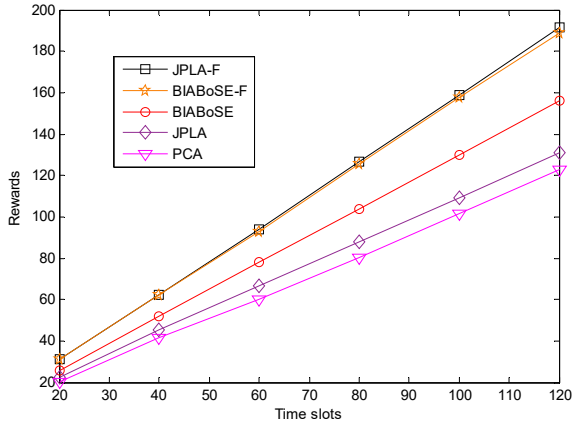

Figure 5. The total rewards versus time slots, λ = 2 and $\omega_t = 1$.

Figure 6 displays the total rewards vary as energy harvesting intensity increases. The system is in a green energy-deficient state, when the energy intensity is low. In this case, the relay can deliver more bits as the intensity increases, which leads to a higher reward. However, the rewards will be constant once the energy intensity reaches a certain threshold, because the battery capacity is limited. It should be noted that the rewards of the BIABoSE are lower than the other algorithms when the green energy is enough due to the discretization of states. However, the BIABoSE achieves better performance in our main application scenario, which is a lack of green energy.

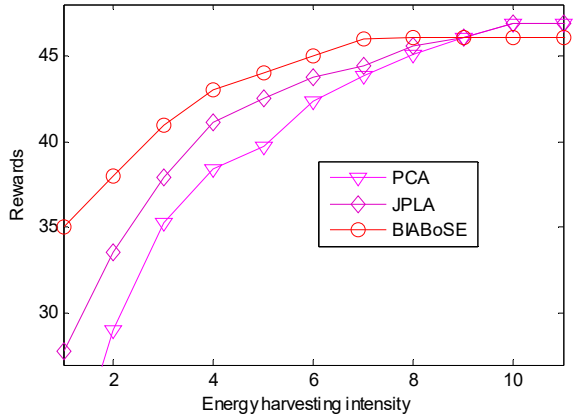

Figure 6. The total rewards versus energy harvesting intensity, N = 30 and $\omega_t = 1$.

4.4. Grid Energy Consumption and Throughput Trade-Off

Figure 7 shows the grid energy consumption and throughput when ω_t takes different values. When ω_t is very small, the energy consumption and throughput of all schemes are similar. In this case, the system has a high demand for grid energy consumption, which will impose strict limits on energy consumption. When ω_t increases, the throughput plays an increasingly important role in the reward. Although our schemes consume a little more energy than the JPLA and PCA, it greatly improves the throughput. When the value of ω_t is large, all schemes pursue maximum throughput regardless of energy consumption costs. Therefore, all throughput gains are very close. However, the BIBAoSE consumes the least energy and is closest to the JPLA-F and BIBAoSE-F. In addition, once the throughput constraints are given, we can find the value of ω_t and get the minimum grid energy consumption.

The energy consumption and throughput are shown in Figure 8. As can be seen from the graph, the BIBAoSE consumes less grid energy than the JPLA when achieves the same throughput. And the BIBAoSE can transmit more bits than the JPLA with the same grid energy supply. In short, the

BIBAoSE has a better trade-off between energy consumption and throughput, which is closer to the ideal situation such as the JPLA-F and BIBAoSE-F.

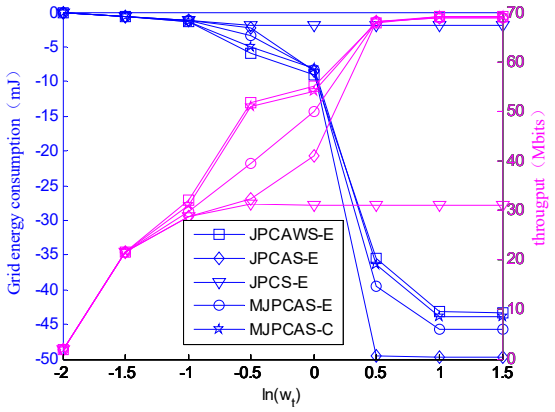

Figure 7. Grid energy consumption and throughput vs. ω_t, N = 30, λ = 2.

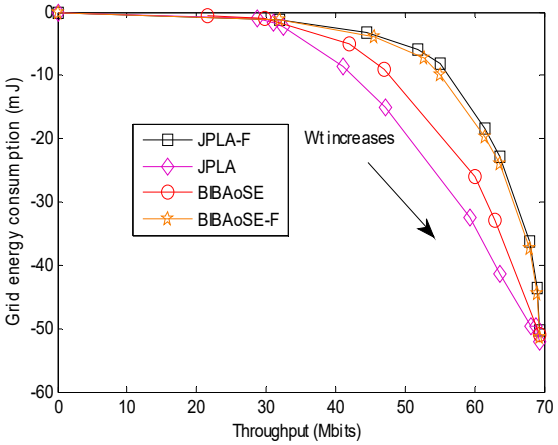

Figure 8. Grid energy consumption versus throughput. N = 30, λ = 2.

5. Conclusions

In this paper, we proposed an online power allocation and link selection strategy to maximize the total rewards of two-hop relay wireless networks where the source node and relay are powered by grid and green energy, respectively. Simulation results show that the total reward of this scheme is optimal under different settings compared with some conventional schemes. Next, we will continue to study energy harvesting technology in multifunctional relay nodes. Then, the research results will be applied to practical scenarios such as 5G heterogeneous networks, the Internet of Things and other networks.

Author Contributions: R.W. and H.X. conceived and designed the experiments; R.W. and H.X. performed the simulations; H.X. and Z.C. wrote the paper; R.W. and L.T. technically reviewed the paper.

Funding: This research was funded by the National Natural Science Foundation of China (No. 51677065).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Liu, T.; Li, J.; Feng, S.; Guan, H.; Yan, S.; Jayakody, D.N.K. On the Incentive Mechanisms for Commercial Edge Caching in 5G Wireless Networks. *IEEE Wirel. Commun.* **2018**, *25*, 72–78. [CrossRef]
2. Ru, W.; Jia, L.; Zhang, G.; Huang, S.; Yuan, M. Energy Efficient Power Allocation for Relay-Aided D2D Communications in 5G Networks. *China Commun.* **2017**, *14*, 54–64.
3. Li, Z.; Fu, X.; Wang, S.; Pei, T.; Li, J. Achievable Rate Maximization for Cognitive Hybrid Satellite-Terrestrial Networks With AF-Relays. *IEEE J. Sel. Areas Commun.* **2018**, *36*, 304–313. [CrossRef]

4. Andrawes, A.; Nordin, R.; Ismail, M. Wireless Energy Harvesting with Cooperative Relaying under the Best Relay Selection Scheme. *Energies* **2019**, *12*, 892. [CrossRef]
5. Lei, C.; Yu, F.R.; Hong, J.; Rong, B.; Li, X.; Leung, V.C.M. Green Full-Duplex Self-Backhaul and Energy Harvesting Small Cell Networks with Massive MIMO. *IEEE J. Sel. Areas Commun.* **2016**, *34*, 3709–3724.
6. Zhu, Z.; Huang, S.; Zheng, C.; Zhou, F.; Zhang, D.; Lee, I. Robust Designs of Beamforming and Power Splitting for Distributed Antenna Systems with Wireless Energy Harvesting. *IEEE Syst. J.* **2018**, *13*, 30–41. [CrossRef]
7. Wang, L.; Wong, K.K.; Shi, J.; Gan, Z.; Robert, W.H., Jr. A New Look at Physical Layer Security, Caching, and Wireless Energy Harvesting for Heterogeneous Ultra-Dense Networks. *IEEE Commun. Mag.* **2017**, *56*, 49–55. [CrossRef]
8. Al-Hraishawi, H.; Baduge, G.A.A. Wireless Energy Harvesting in Cognitive Massive MIMO Systems with Underlay Spectrum Sharing. *IEEE Wirel. Commun. Lett.* **2017**, *6*, 134–137. [CrossRef]
9. Zhao, C.; Cai, L.X.; Yu, C.; Shan, H. Sustainable Cooperative Communication in Wireless Powered Networks with Energy Harvesting Relay. *IEEE Trans. Wirel. Commun.* **2017**, *16*, 8175–8189.
10. Ozel, O.; Tutuncuoglu, K.; Yang, J.; Ulukus, S.; Yener, A. Transmission with Energy Harvesting Nodes in Fading Wireless Channels: Optimal Policies. *IEEE J. Sel. Areas Commun.* **2011**, *29*, 1732–1743. [CrossRef]
11. Zhao, M.; Zhao, J.; Zhou, W.; Zhu, J.; Zhang, S. Energy efficiency optimization in relay-assisted networks with energy harvesting relay constraints. *China Commun.* **2015**, *12*, 84–94. [CrossRef]
12. Ahmed, I.; Ikhlef, A.; Schober, R.; Mallik, R.K. Power Allocation for Conventional and Buffer-Aided Link Adaptive Relaying Systems with Energy Harvesting Nodes. *IEEE Trans. Wirel. Commun.* **2014**, *13*, 1182–1195. [CrossRef]
13. Zhi, C.; Dong, Y.; Fan, P.; Letaief, K.B. Optimal Throughput for Two-Way Relaying: Energy Harvesting and Energy Co-Operation. *IEEE J. Sel. Areas Commun.* **2016**, *34*, 1448–1462.
14. Luo, Y.; Zhang, J.; Letaief, K.B. Relay selection for energy harvesting cooperative communication systems. In Proceedings of the IEEE Global Communications Conference, Atlanta, GA, USA, 9–13 December 2013.
15. Lee, Y.H.; Liu, K.H. Battery-aware relay selection for energy-harvesting relays with energy storage. In Proceedings of the IEEE 26th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC), Hong Kong, China, 30 August–2 September 2015.
16. Wang, F.; Guo, S.; Yang, Y.; Xiao, B. Relay Selection and Power Allocation for Cooperative Communication Networks with Energy Harvesting. *IEEE Syst. J.* **2016**, *12*, 1–12. [CrossRef]
17. Yuan, W.; Li, P.Q.; Liang, H.; Shen, X. Optimal Relay Selection and Power Control for Energy-Harvesting Wireless Relay Networks. *IEEE Trans. Green Commun. Netw.* **2018**, *2*, 471–481.
18. Yu, P.S.; Lee, J.; Quek, T.Q.S.; Hong, Y.-W.P. Traffic Offloading in Heterogeneous Networks with Energy Harvesting Personal Cells—Network Throughput and Energy Efficiency. *IEEE Trans. Wirel. Commun.* **2015**, *15*, 1146–1161. [CrossRef]
19. Zhang, S.; Zhang, N.; Zhou, S.; Gong, J.; Niu, Z.; Shen, X. Energy-Aware Traffic Offloading for Green Heterogeneous Networks. *IEEE J. Sel. Areas Commun.* **2016**, *34*, 1116–1129.
20. Dhillon, H.S.; Li, Y.; Nuggehalli, P.; Pi, Z.; Andrews, J.G. Fundamentals of Heterogeneous Cellular Networks with Energy Harvesting. *IEEE Trans. Wirel. Commun.* **2014**, *13*, 2782–2797.
21. Ahmed, I.; Ikhlef, A.; Schober, R.; Mallik, R.K. Joint Power Allocation and Relay Selection in Energy Harvesting AF Relay Systems. *IEEE Wirel. Commun. Lett.* **2013**, *2*, 239–242. [CrossRef]
22. Huang, C.; Zhang, R.; Cui, S. Throughput Maximization for the Gaussian Relay Channel with Energy Harvesting Constraints. *IEEE J. Sel. Areas Commun.* **2013**, *31*, 1469–1479. [CrossRef]
23. Yu, G.; Jiang, Y.; Xu, L.; Li, G.Y. Multi-Objective Energy-Efficient Resource Allocation for Multi-RAT Heterogeneous Networks. *IEEE J. Sel. Areas Commun.* **2015**, *33*, 2118–2127. [CrossRef]
24. Gong, J.; Zhou, Z.; Zhou, S. On the Time Scales of Energy Arrival and Channel Fading in Energy Harvesting Communications. *IEEE Trans. Green Commun. Netw.* **2018**, *2*, 482–492. [CrossRef]
25. Benjaafar, S.; Morin, T.L.; Talavage, J.J. The strategic value of flexibility in sequential decision making. *Eur. J. Oper. Res.* **1995**, *82*, 438–457. [CrossRef]
26. Bertsekas, D.P. *Dynamic Programming and Optimal Control*; Athena Sci.: Belmont, MA, USA, 2005.
27. Hu, Q.; Chen, X. The finiteness of the reward function and the optimal value function in Markov decision processes. *Math. Methods Oper. Res.* **1999**, *49*, 255–266.

28. Mao, Y.; Zhang, J.; Letaief, K.B. Grid Energy Consumption and QoS Tradeoff in Hybrid Energy Supply Wireless Networks. *IEEE Trans. Wirel. Commun.* **2016**, *15*, 3573–3586. [CrossRef]
29. Song, H.; Liu, C.C.; Lawarrée, J.; Dahlgren, R.W. Optimal electricity supply bidding by Markov decision process. *IEEE Trans. Power Syst.* **2000**, *15*, 618–624. [CrossRef]