RESEARCH

MDP-based handover policy in wireless relay systems

Xiaoyu Dang[1], Jin-Yuan Wang[2*] and Zhe Cao[1]

Abstract

Wireless relay transmission has been considered as a supplementary technology for future wireless communication system, and handover is a key element in wireless relay transmissions to support users' mobility. Handover failure will result in the forced termination of an ongoing call, and a good handover scheme will provide good system performance. This article focuses on the handover decision problem in wireless relay transmission systems. Initially, the architecture of a single-cell relay system is introduced. The handover decision problem is formulated as a Markov decision process (MDP). In the proposed model, a profit function is used to evaluate the quality-of-service of the chosen serving node, and a cost function is used to model processing load and signaling cost. Based on the profit and cost functions, the reward function is formulated. Moreover, the objective is to maximize the expected total reward per connection. The value iteration algorithm is employed to determine the optimal handover policy. Furthermore, the analysis is generalized to multi-cell relay systems. Numerical results show that the proposed MDP-based handover policy is well behaved in wireless relay systems.

Keywords: Wireless relay systems, Handover policy, Markov decision process, Expected total reward

1 Introduction

Recently, wireless communications have been developing very rapidly from code division multiple access to orthogonal frequency division multiple access (OFDMA) techniques [1-3]. As the fast growing of users' demands, the new generation wireless communication systems will provide high data rate and wide coverage. However, due to the restriction of scarce spectrum resources, the spectral efficiency of the new generation wireless communication systems is very limited [4]. The relay transmission, which deploys multiple relay nodes between a source node and a destination node, has attracted much attention in recent years. Due to its great potential to improve spectral efficiency and coverage area, relay transmission has been considered as a supplementary technology for the new generation wireless communication systems. As a result, the IEEE 802.11 Task Group and 802.16 Task Group are actively working on the standardization of relay transmission protocols [5,6].

The basic idea of wireless relay transmission was introduced by van der Meulen [7], and was comprehensively studied by Cover and El Gamal [8]. So far, wireless relay transmission has been investigated in various aspects, including information-theoretic capacity [9,10], diversity [11], outage performance [12], network coding [13], and power allocation [14,15]. It should be noted that handover [16,17], which provides the seamless mobility, is also an important issue in wireless relay systems. In conventional cellular systems, handover occurs only when a mobile station (MS) moves to different cells or different sectors of the same cell. However, the introduction of relay stations (RSs) in cellular systems creates additional handover scenarios. Figure 1 depicts diverse handover scenarios in wireless relay systems [18]. In scenario 1, an MS performs handover between two different RSs in the same cell. In scenario 2, an MS changes its communication node from a base station (BS) to the RS of the same cell, or vice versa. Scenario 3 is exactly the same as the inter-BS handover in the conventional cellular systems. In scenario 4, an MS performs handover from an RS to the RS of different cells. Scenario 5 is the process of an MS moves from a BS to the RS of different cells. Each mobile connection may experience a number of handovers during its lifetime

*Correspondence: jinyuan798@seu.edu.cn
[2]National Mobile Communications Research Laboratory, Southeast University, Nanjing 210096, China
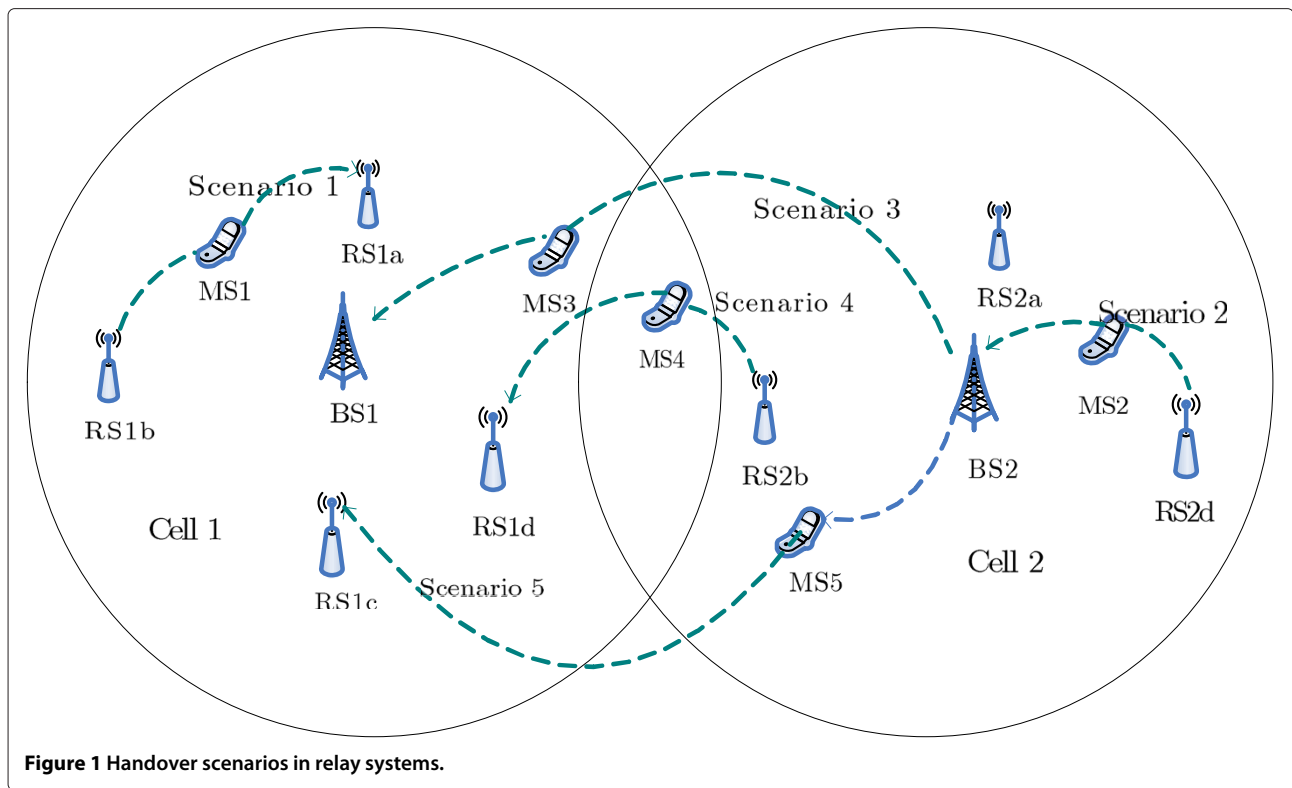Full list of author information is available at the end of the article

Figure 1 Handover scenarios in relay systems.

connection. An efficient handover decision policy is beneficial to the system performance. Therefore, the handover is an important problem in wireless relay transmission systems. Recently, the handover schemes in wireless relay transmission systems have been defined in the base line document for the draft standard of IEEE 802.16j [6].

The handover policies for wireless communications have been reported in some literatures. Stemm and Katz [19] proposed a vertical handover scheme, where the handover decisions only depend on the presence or absence of beacon packets. A series of researches about handover based on the received signal strength have been carried out. In [20], the handover algorithms based on the received signal strength were reviewed, and advanced techniques (such as hypothesis testing, dynamic programming, and pattern recognition based on neural networks or fuzzy logic) were mentioned. An adaptive timer handoff (ATHO) algorithm was proposed in [21], which was based on the received signal strength with a hysteresis timer to adapt the ping-pong effect in mobile communication systems. In [22], a practical approach based on GSM measurement data was employed, and a handover algorithm was proposed to improve the handover performance. Jiang et al. [23] proposed a novel scheme using uplink and downlink signals for intra multi-hop relay BS handover with transparent RS. In [24], an optimization handover scheme for mobile users was proposed

to minimize the delay and maximize the throughput by employing a dwell-timer. Although all above works have investigated the handover problems in wireless systems, the link quality conditions are only taken as the handover decision criteria, and the quality-of-service (QoS) and the processing load during handovers are not considered. To the best of the authors' knowledge, few previous studies take the QoS and processing load together as the handover criteria in wireless relay systems.

In this article, we focus this study on a handover policy in wireless relay systems. At first, a relay transmission system with a single cell is introduced. In order to obtain better system performance, the adaptive modulation and coding schemes are employed. Furthermore, the available transmission rates for both the direct transmission and relay transmission links have been analyzed. Due to the mobility of MSs, a handover decision problem is then considered, which is formulated as a Markov decision process (MDP). After that, a profit function is used to evaluate the QoS of the chosen serving node, which represents the benefit that the MS can gain by choosing a serving node (i.e., the BS or some RS). The cost function is also considered, which captures the processing and signal load incurred when the connection switch the current serving node to another. Based on the profit and cost functions, the reward function is formulated. The objective of the handover problem is to maximize the expected total

reward per connection. An optimal MDP-based handover decision policy is proposed by employing the value iteration algorithm (VIA). Furthermore, the proposed MDP-based handover policy is generalized to multi-cell relay systems.

The remainder of this article is organized as follows. The system model is described in Section 2. Section 3 formulates the handover decision process as an MDP, after that the optimality equations and the VIA are proposed. Extensions to multiple cell systems are discussed in Section 4. Numerical results are presented in Section 5 before conclusions are drawn in Section 6.

2 System model

Consider the downlink transmission of a single-cell relay system, which is composed of one BS, M RSs, and multiple MSs, as shown in Figure 2. The BS can communicate with MSs directly or with the help of one RS. Depending on whether the BS communicates with MSs directly or with the help of RS, the transmissions can be classified into direct transmission and relay transmission. Moreover, the transmit power of BS (RS) is equally distributed to its serving MSs. Here, each MS is assumed to operate in a pre-assigned orthogonal channel, such as non-overlapping time/frequency slot.

This assumption holds for most practical systems such as time division multiple access, frequency division multiple access (FDMA), and OFDMA systems [25]. Therefore, no inter-channel interferences exist among all MSs. In order to maximize system performance, the adaptive modulation and coding schemes are employed. Assuming that the BS or RSs can transmit data to the MS with $W - 1$ nonzero fixed transmission rates, which can be denoted as $R^{(1)}, \ldots, R^{(W-1)}$. Without loss of generality, we suppose that $R^{(W-1)} > \cdots > R^{(1)} > 0$. If zero is regarded as one special transmission rate, the set of downlink transmission rates can be denoted as $\Lambda = \left\{ R^{(0)}, R^{(1)}, \ldots, R^{(W-1)} \right\}$, where $R^{(0)} = 0$. To guarantee the receiving accuracy of certain downlink transmission rate, it is expected that the received signal-to-noise ratio (SNR) must be no less than the target SNR. For a downlink transmission rate $R^{(\omega)}(\omega = 0, 1, \ldots, W - 1)$, the corresponding target SNR is denoted by $\gamma^{(\omega)}$. Obviously, these target SNRs should satisfy

$$0 = \gamma^{(0)} < \gamma^{(1)} < \cdots < \gamma^{(W-1)} < \gamma^{(W)} \triangleq \infty \qquad (1)$$

When the received SNR $\gamma \in \left[\gamma^{(\omega)}, \gamma^{(\omega+1)} \right)$, the transmission rate can be determined by $R^{(\omega)}$ at most.
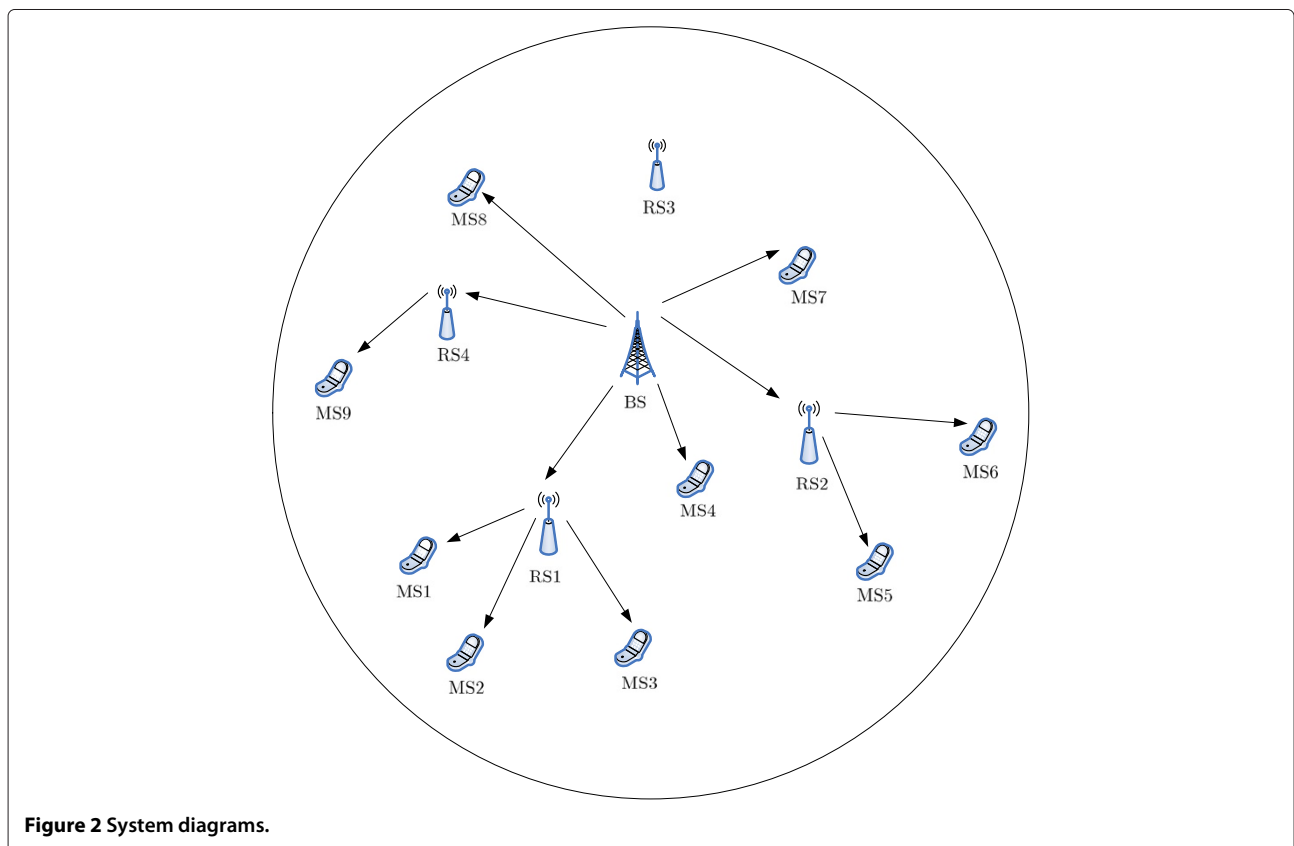

Figure 2 System diagrams.

Mathematically, given a received SNR γ, the transmission rate r can be obtained by

$$r = \xi(\gamma) \triangleq \sum_{\omega=0}^{W-1} R^{(\omega)} \left[U\left(\gamma - \gamma^{(\omega)}\right) - U\left(\gamma - \gamma^{(\omega+1)}\right) \right] \tag{2}$$

where $U(t - t_0)$ is the Heaviside unit step function defined by

$$U(t - t_0) = \begin{cases} 1, & \text{for } t \geq t_0 \\ 0, & \text{for } t < t_0 \end{cases} \tag{3}$$

Note that the relationship between the received SNR γ and the transmission rate r can be seen in Figure 3.

For the direct transmission case, an MS can receive signal from the BS directly. The received downlink signal y_0 can be expressed as

$$y_0 = \sqrt{\frac{P_0}{N_0}} G_0 x_0 + z_0 \tag{4}$$

where P_0 denotes the total transmit power of the BS, N_0 is the number of MSs served by the BS, x_0 is the transmitted information symbol with unit energy at the BS, G_0 is the channel gain between the BS and the MS, and z_0 is the additive white Gaussian noise. Without loss of generality, we assume that the noise level is the same for all links and

is denoted by σ^2. Then, the directly received SNR γ_0 can be expressed as

$$\gamma_0 = \frac{P_0 G_0}{N_0 \sigma^2} \tag{5}$$

Therefore, in this case, the maximum transmission rate r_0 can be obtained by

$$r_0 = \xi(\gamma_0) \tag{6}$$

For the relay transmission case, we concentrate on the BS–RSm–MS link. In this case, the transmission consists of two phases. In phase 1, the BS first transmits the information to RS m, which is selected to relay the information. Therefore, the received signals at RS m can be expressed as

$$y_m^{(1)} = \sqrt{\frac{P_0}{N_0^{(1)}}} G_m^{(1)} x_m^{(1)} + z_m^{(1)} \tag{7}$$

where $N_0^{(1)}$ is the number of the RSs served by the BS in phase 1, $G_m^{(1)}$ is the channel gain between the BS and RS m, $z_m^{(1)}$ is the received additive white Gaussian noise. Therefore, the received SNR at RS m can be derived as

$$\gamma_m^{(1)} = \frac{P_0 G_m^{(1)}}{N_0^{(1)} \sigma^2} \tag{8}$$

Therefore, the transmission rate $r_m^{(1)}$ on the link between the BS and RS m can be obtained by

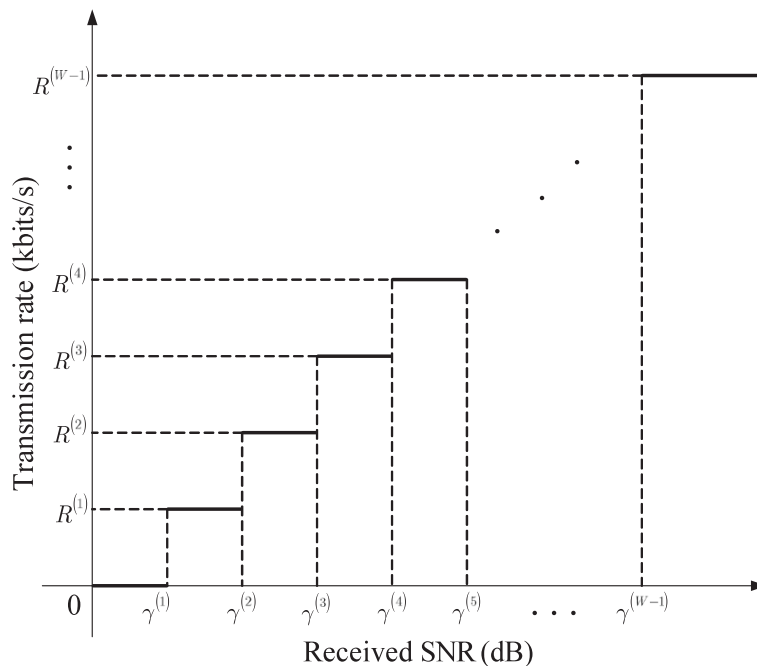$$r_m^{(1)} = \xi\left(\gamma_m^{(1)}\right) \tag{9}$$

Figure 3 The transmission rate versus the received SNR.

In phase 2, amplify-and-forward relay transmission strategy is employed [26], that is, RS m amplifies $y_m^{(1)}$ with transmission power $P_m \big/ N_m^{(2)}$ and forwards it to the MS, where P_m is the total transmit power of RS m and $N_m^{(2)}$ is the numbers of MSs served by RS m in phase 2. Accordingly, the received signal of the MS can be expressed as

$$y_m^{(2)} = \sqrt{\frac{P_m}{N_m^{(2)}}} G_m^{(2)} x_m^{(2)} + z_m^{(2)} \qquad (10)$$

where

$$x_m^{(2)} = \frac{y_m^{(1)}}{\left| y_m^{(1)} \right|} \qquad (11)$$

is the unit-energy transmitted signal that RS m received from the BS in phase 1, $G_m^{(2)}$ is the channel gain between RS m and the MS, and $z_m^{(2)}$ is the received noise at the MS. Then, in this phase, the received SNR at the MS can be expressed as

$$\gamma_m^{(2)} = \frac{P_m G_m^{(2)}}{N_m^{(2)} \sigma^2} \qquad (12)$$

Accordingly, the transmission rate $r_m^{(2)}$ in phase 2 can be obtained by

$$r_m^{(2)} = \xi \left(\gamma_m^{(2)} \right) \qquad (13)$$

Finally, the available transmission rate on the BS−RSm−MS link can be determined as

$$r_m = \frac{1}{2} \min \left(r_m^{(1)}, r_m^{(2)} \right) \qquad (14)$$

where the coefficient $1/2$ is due to the fact that cooperative transmission uses half of the resources (e.g., time slots, frequency bands, orthogonal codes).

Due to the mobility, each MS will monitor the whole system all the time and has the choices of communicating directly with the BS or with the help of one RS, which involves the handover decision problems. The following section will solve this handover problem using MDP.

3 MDP-based handover decision

MDP, also referred to as stochastic dynamic programs or stochastic control problems, are models for sequential decision making when outcomes are uncertain. The MDP model consists of decision epochs, states, actions, rewards, and transition probabilities [27]. Choosing an action in a state generates a reward and determines the state at the next decision epoch through a transition probability function. Policies or strategies are prescriptions of which action to choose under any eventuality at every future decision epoch. Decision makers seek policies which are optimal in some sense. MDP has already been successfully used to solve a variety of problems, including finance [28], admission control [29], mobility-related issues in the areas of mobile communications, and in wireless sensor networks [30,31]. Therefore, MDP is a promising candidate in solving the handover control problems in wireless relay systems.

In the following, we will describe how the handover decision problem is formulated as an MDP. Then, the optimality equations and the VIA are introduced. To facilitate the description, in the following, the BS and all RSs are also called as serving nodes. Specially, the serving node 0 refers to the BS and the serving node m ($m = 1, \ldots, M$) refers to RS m.

3.1 Decision epoch, action, and state

Referring to Figure 4, the sequence $T = \{1, 2, \ldots, Q\}$ represents the time of successive decision epochs, and the random variable Q denotes the time that the connection terminates. At each decision epoch, the MS has to decide whether the connection should use the current serving node or choose another serving node based on its current state. In this article, the state space of an MS is denoted by S, and the number of states that an MS can possibly be in is finite. The state of the MS contains information such as the current serving node that the MS connects to, and the available transmission rates that all serving nodes offer. Specifically, the state space can be expressed as follows

$$S = \{0, 1, 2, \ldots, M\} \times R_0 \times R_1 \times R_2 \times \cdots \times R_M \quad (15)$$

where "$\times$" denotes the Cartesian product, $\{0, 1, 2, \ldots, M\}$ denotes the set of current serving nodes that the MS connects to, R_0 denotes the set of the available transmission rates on the direct transmission link, and R_m ($m = 1, \ldots, M$) denotes the set of the available transmission rates on the BS−RSm−MS link. Obviously, it can be known that $R_0 = \Lambda$. Furthermore, according to Equation (14), R_m can be derived as

$$R_m = \{r | 2r \in \Lambda\}, \quad \forall m \in \{1, 2, \ldots, M\} \qquad (16)$$

Based on the current state, the MS has to choose an action at each decision epoch. Considering the MS can only select a serving node from the BS and all RSs, the action set can be defined as $A \triangleq \{0, 1, 2, \ldots, M\}$, where A consists of all the serving nodes that MS can handover to.
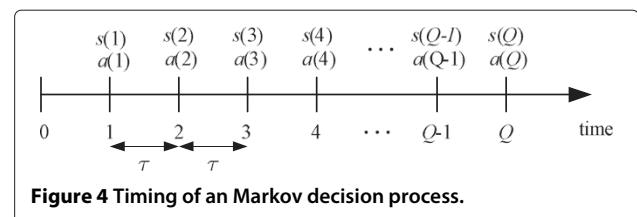

Figure 4 Timing of an Markov decision process.

3.2 Transition probability

Let vector $s(t) = [i, r_0, r_1, \ldots, r_M]$ denote the state of the MS at the tth decision epoch, where i denotes the current chosen serving node, r_0 denotes the current available transmission rate on the direct transmission link, r_m ($m = 1, \ldots, M$) denotes the current available transmission rate on the BS–RSm–MS link. Based on the current state $s(t)$, if the chosen action is $a(t) = j$, the probability transition function that the next state $s(t+1) = [j, r'_0, r'_1, \ldots r'_M]$ is given by

$$P[s(t+1)|s(t),a(t)] = \begin{cases} \prod_{m=0}^{M} P[r'_m | r_m, a(t)], & \text{if } a(t) = j \\ 0, & \text{otherwise} \end{cases},$$

(17)

where $P[r'_m | r_m, a(t)]$ denotes the transition probability of the transmission rate of the serving node m. Although all serving nodes are collected in the system, the transmission rate of each serving node is only rely on the number of users supported by the serving node and the channel gain between the node and the MS. Therefore, the transition probability of the transmission rate of each serving node in Equation (17) is assumed to be independent with each other. Specially, if the values of the transmission rates are guaranteed for the duration of the connection, the transition probability $P[r'_m | r_m, a(t)]$ in this special case can be expressed as

$$P[r'_m | r_m, a(t)] = \begin{cases} 1, & \text{if } r'_m = r_m \\ 0, & \text{otherwise} \end{cases}$$

(18)

Note that the functions (17) and (18) are Markovian because the state transition probability only depends on the current state and action but not on the previous states.

3.3 Reward function

Let $z(s(t), a(t))$ denote the reward that the MS receives after the tth decision epoch. In this section, the reward function is defined as

$$z(s(t), a(t)) = f(s(t), a(t)) - g(s(t), a(t))$$ (19)

where $f(s(t), a(t))$ is the profit function and reflects the QoS provided by the chosen serving node at the tth decision epoch. $g(s(t), a(t))$ denotes the cost function, which captures the processing and signal load incurred when the connection switch the current serving node to another.

Given the current state $s(t) = [i, r_0, r_1, \ldots, r_M]$, where i denotes the current serving node used by the connection, r_m denotes the available rate provided by serving node m. At each decision epoch, based on the current state $s(t)$, the MS should choose

an action $a(t) \in A$. Therefore, the profit function can be defined as

$$f(s(t), a(t)) = \begin{cases} 1, & \text{if } r_i = \max_{k \in \{0,1,\ldots,M\}} \{r_k\}, \ a(t) = i, \\ 0, & \text{if } r_i = \max_{k \in \{0,1,\ldots,M\}} \{r_k\}, \ a(t) \neq i, \\ \dfrac{r_{a(t)} - r_i}{\max_{k \in \{0,1,\ldots,M\}} \{r_k - r_i\}}, & \text{if } r_i \neq \max_{k \in \{0,1,\ldots,M\}} \{r_k\}, \ r_{a(t)} > r_i \\ 0, & \text{if } r_i \neq \max_{k \in \{0,1,\ldots,M\}} \{r_k\}, \ r_{a(t)} \leq r_i \end{cases}$$

(20)

The profit function (in terms of transmission rate) can be assessed as follows. Given that the MS is currently connecting to serving node i, if serving node i is the one which supports the highest rate among others and the chosen action $a(t) = i$, the profit is set to be 1, otherwise, the profit is set to be 0. However, when serving node i is not the one which supports the highest rate, the profit that it can obtain is represented by a fraction, in which the numerator is the MS's actual increase of rate by choosing action $a(t)$ in state $s(t)$, and the denominator is the MS's maximum possible increase of rate.

The cost function $g(s(t), a(t))$ is defined as

$$g(s(t), a(t)) = \begin{cases} K_{i,a(t)}, & \text{if } a(t) \neq i \\ 0, & \text{if } a(t) = i \end{cases}$$

(21)

It can be known from Equation (21) that, when the new serving node $a(t)$ is the same as the current serving node i (i.e., $a(t) = i$), there is no handover happens, and thus the cost is set to be zero. However, when the new serving node $a(t)$ is different from the current serving node i (i.e., $a(t) \neq i$), the handover will happen, and the cost in this case is set to be $K_{i,a(t)}$. Note that $K_{i,a(t)}$ denotes the switching cost from the current serving node i to the new serving node $a(t)$. The value of $K_{i,a(t)}$ depends on several factors, such as the types of handover and the current traffic load on the serving node.

3.4 Expected total reward

A decision rule prescribes a procedure for action selection in each state at a specified decision epoch. Deterministic Markov decision rules are functions $\delta(t) : S \to A$, which specify the action choice $a(t)$ when the system occupies state $s(t)$ at the tth decision epoch. A policy $\pi = (\delta(1), \delta(2), \ldots, \delta(Q))$ is a sequence of decision rules to be used at all decision epochs.

Let $v^\pi(s(0))$ denote the expected total reward from the first decision epoch until the handover decision period elapses, given that the policy π is used with an initial state $s(0)$. We have

$$v^\pi(s(0)) = E^\pi_{s(0)} \left[E_Q \left\{ \sum_{t=1}^{Q} z(s(t), a(t)) \right\} \right]$$

(22)

where $E_{s(0)}^{\pi}$ denotes the expectation with respect to policy π and initial state $s(0)$, and E_Q denotes the expectation with respect to random variable Q. Note that a different policy π and initial state $s(0)$ will change the chosen action $a(t)$. This will also cause a different transition probability function $P[s(t+1)|s(t),a(t)]$ to be used in the expectation $E_{s(0)}^{\pi}$. The random variable Q, which denotes the connection termination time, is assumed to be geometrically distributed with mean $1/(1-\lambda)$ [32]. That is

$$P(Q = q) = \lambda^{q-1}(1-\lambda), \ \forall n = 1, 2, 3, \ldots \quad (23)$$

Therefore, Equation (22) can be further written as

$$v^{\pi}(s(0)) = E_{s(0)}^{\pi} \left\{ \sum_{q=1}^{\infty} \sum_{t=1}^{q} z(s(t), a(t)) \lambda^{q-1}(1-\lambda) \right\} \quad (24)$$

Since $\sum_{q=1}^{\infty} \sum_{t=1}^{q} (\cdot) = \sum_{t=1}^{\infty} \sum_{q=t}^{\infty} (\cdot)$, by interchanging the order of the summation, we have

$$
\begin{aligned}
v^{\pi}(s(0)) &= E_{s(0)}^{\pi} \left\{ \sum_{t=1}^{\infty} \sum_{q=t}^{\infty} z(s(t), a(t)) \lambda^{q-1}(1-\lambda) \right\} \\
&= E_{s(0)}^{\pi} \left\{ \sum_{t=1}^{\infty} \lambda^{t-1} z(s(t), a(t)) \right\}
\end{aligned}
\quad (25)
$$

where $\lambda \in [0, 1)$ is the discount factor of the model.

Our optimization problem is to maximize the expected total discounted reward. The π^* is defined to be the optimal policy in Π if $v^{\pi^*} \geq v^{\pi}$ for all $\pi \in \Pi$. A policy is said to be stationary if $\delta(t) = \delta$ for all t. A stationary policy has the form $\pi = (\delta, \delta, \ldots, \delta)$. For convenience, π is denoted as δ. Our objective is to determine an optimal stationary deterministic policy δ^*, which maximizes the expected total discounted reward given by Equation (25).

3.5 Optimality equations and the VIA

In this section, the optimality equations which maximize the expected total reward are proposed, and following that, a VIA is used to determine a stationary optimal handover policy.

Let $v(s)$ denote the maximum expected total reward, given the initial state s. That is

$$v(s) = \max_{\pi \in \Pi} v^{\pi}(s) \quad (26)$$

Referring to [27], the optimality equations can be written as

$$v(s) = \max_{a \in A} \left\{ z(s,a) + \sum_{s' \in S} \lambda P[s'|s,a] v(s') \right\} \quad (27)$$

where s' denotes the next state while we select action a in state s. The solutions of the optimality equations

correspond to the maximum expected total reward $v(s)$ and the MDP optimal policy $\delta^*(s)$. Note that the MDP optimal policy $\delta^*(s)$ indicates the decision as to which node to choose from, given that the current state is s.

The VIA [27,32] is employed in this article to determine a stationary deterministic optimal policy and the corresponding expected total reward. The stepwise procedures of the VIA algorithm are described as follows:

Step (1): Set $v^0(s) = 0$ for each state s. Specify $\varepsilon > 0$ and set $k = 0$.
Step (2): For each state s, compute $v^{k+1}(s)$ by using

$$v^{k+1}(s) \max_{a \in A} \left\{ z(s,a) + \sum_{s' \in S} \lambda P[s'|s,a] v^k(s') \right\}$$

Step (3): If $\|v^{k+1} - v^k\| < \epsilon(1-\lambda)/2\lambda$ then
Go to Step (4).
Else
Let $k = k + 1$, and return to Step (2).
EndIf
Step (4): For each state $s \in S$, compute the stationary optimal policy by using

$$\delta(s) = argmax \left\{ z(s,a) + \sum_{s' \in S} \lambda P[s'|s,a] v^{k+1}(s') \right\}$$

Step (5): Stop.

In the VIA, the normal function $\|\cdot\|$ is defined as $\|v\| = \max |v(s)|$ for $s \in S$. Note that the VIA is convergent because the operation in Step (2) corresponds to a contraction mapping. Each iteration of the VIA is performed in $O(|A||S|^2)$, i.e., the convergence rate of the VIA is linear.

A stationary deterministic optimal policy table can be created according to this algorithm. The MSs are assumed to periodically receive information from the serving nodes. The advertised information from each serving node may include, among other parameters, the achieved transmission rate and the handover cost. In each time of period, the MSs decide whether the connection should use the current serving node or reroute to another serving node according to the stationary deterministic optimal policy table.

4 Extensions to multiple cell systems

The proposed MDP-based handover policy can be generalized to multi-cell systems. Assuming that there are E cells in the system, and each cell has one BS and N RSs. Let us define B_e ($e = 1, 2, \ldots, E$) as the BS in cell e, and define $R_n^{(e)}$ ($e = 1, 2, \ldots, E$, $n = 1, 2, \ldots, N$) as the RS n in cell e. Then, the directly received signal-to-interference-plus noise ratio (SINR) provided by BS e can be expressed as

$$\gamma_{B_e,M} = \frac{P_{B_e} G_{B_e,M}}{N_{B_e}(\sigma^2 + I_{B_e,M})} \quad (28)$$

where P_{B_e} is the total transmit power of BS e, N_{B_e} is the number of nodes (i.e., MSs in direct transmission case) served by BS e, and $G_{B_e,M}$ is the channel gain from BS e to the MS, $I_{B_e,M}$ denotes the inter-cell interference (ICI). Therefore, in this case, the actual available maximum direct transmission rate $r_{B_e,M}, (e = 1, 2, \ldots, E)$ can be obtained by

$$r_{B_e,M} = \xi\left(\gamma_{B_e,M}\right) \tag{29}$$

For the relay transmission case, the SINR at RS n of cell e in phase 1 can be derived as

$$\gamma^{(1)}_{B_e,R_n^{(e)}} = \frac{P_{B_e} G_{B_e,R_n^{(e)}}}{N_{B_e}\left(\sigma^2 + I^{(1)}_{B_e,R_n^{(e)}}\right)} \tag{30}$$

where N_{B_e} is the number of nodes (i.e., RSs in phase 1 of the relay transmission case) served by BS e, $G_{B_e,R_n^{(e)}}$ is the channel gain between the BS and RS n in cell e, and $I^{(1)}_{B_e,R_n^{(e)}}$ is the ICI. Then, the transmission rate $r^{(1)}_{B_e,R_n^{(e)}}$ on the BS-$R_n^{(e)}$ link in cell e can be obtained by

$$r^{(1)}_{B_e,R_n^{(e)}} = \xi\left(\gamma^{(1)}_{B_e,R_n^{(e)}}\right) \tag{31}$$

Similarly, in phase 2, the received SINR on the $R_n^{(e)}$-MS link in cell e can be expressed as

$$\gamma^{(2)}_{B_e,R_n^{(e)}} = \frac{P_{R_n^{(e)}} G_{R_n^{(e)},M}}{N_{R_n^{(e)}}\left(\sigma^2 + I^{(2)}_{B_e,R_n^{(e)}}\right)} \tag{32}$$

where $P_{R_n^{(e)}}$ is the total transmit power of RS n in cell e, $N_{R_n^{(e)}}$ is the number of MSs served by RS n in cell e, $G_{R_n^{(e)},M}$ and $I^{(2)}_{B_e,R_n^{(e)}}$ are the channel gain and ICI from RS n to the MS in cell e. Then, the transmission rate $r^{(2)}_{B_e,R_n^{(e)}}$ on the $R_n^{(e)}$-MS link in cell e can be obtained by

$$r^{(2)}_{B_e,R_n^{(e)}} = \xi\left(\gamma^{(2)}_{B_e,R_n^{(e)}}\right) \tag{33}$$

Therefore, the available rate of the MS on the $BS_e-R_n^{(e)}-$MS link can be expressed as

$$r_{B_e,R_n^{(e)}} = \frac{1}{2}\min\left(r^{(1)}_{B_e,R_n^{(e)}}, r^{(2)}_{B_e,R_n^{(e)}}\right) \tag{34}$$

According to MDP, the action set can be defined as $A \triangleq \left\{B_e, R_n^{(e)} (e = 1, \ldots, E, n = 1, \ldots, N)\right\}$, and the state space can be expressed as

Table 1 The SNR threshold of different modulation and coding schemes maintaining the target BER of 10^{-3}

Modulation and coding scheme	Target SNR (dB)
1/2-BPSK	3
1/2-QPSK	4.5
2/3-QPSK	6.5
3/4-QPSK	8
5/6-QPSK	9.3
1/2-16QAM	11.5
2/3-16QAM	13

$$S = \left\{B_1, R_1^{(1)}, \ldots, R_N^{(1)}, \ldots, B_E, R_1^{(E)}, \ldots, R_N^{(E)}\right\}$$
$$\times R_{B_1} \times R_{R_1^{(1)}} \times \cdots \times R_{R_N^{(1)}} \times \cdots \times R_{B_E} \times R_{R_1^{(E)}} \cdots \times R_{R_N^{(E)}} \tag{35}$$

where $R_{B_e} (e = 1, 2, \ldots, E)$ denotes the set of the available transmission rate of the BS in cell e, and $R_{R_n^{(e)}} (e = 1, 2, \ldots, E, n = 1, 2 \ldots, N)$ denotes the set of the available transmission rate of RS n in cell e.

Similar to the single-cell system, the handover decision policy can be determined by VIA. Note that, the introduction of multiple-cell systems will result in the increase of the state space of MDP. Obviously, computational complexity will increase accordingly.

5 Numerical results

In order to evaluate the performance, it is desired to compare our proposed MDP-based handover policy with other policies through computer simulations. Unfortunately, to the best of the authors' knowledge, few previous studies take the QoS and processing load together as the handover criteria in wireless relay systems. Here, a nearest distance handover policy and a biggest channel gain handover policy are taken as the comparative policies. In the nearest distance handover policy, each MS chooses the nearest serving node at each decision epoch. In the biggest channel gain handover policy, the serving node to be selected in each decision epoch is the one that has the biggest channel gain. Then, under the same simulation environments, the numerical results of the proposed MDP-based handover policy are compared with the nearest distance handover policy and the biggest channel gain handover policy.

Table 2 The parameters of the relay system

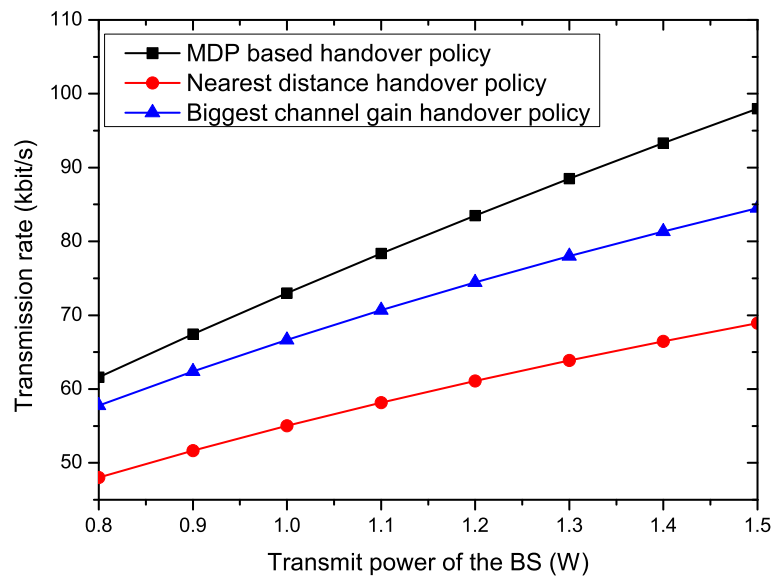| Parameters	Value
ε	10^{-3}
λ	0.9
α	1
$K_{i,a(t)} = K_{a(t),i}, a(t) \neq i$	0.75

Figure 5 The system transmission rates with different transmit powers of BS.

A single-cell wireless relay system is used as the test system. In the system, all serving nodes and MSs are located in a square area with 2000-m side length. Specifically, the BS is located in the center of the cell and its coordinate is set to be (0 m, 0 m). To simplify the dimension of the state space, four RSs are configured in the cell, which are located at (600 m, 0 m), (0 m, 600 m), (−600 m, 0 m), and (0 m, −600 m), respectively. The initial positions of all MSs are randomly generated in the cell. A mobility pattern is employed for each MS, i.e., the speed of each MS follows uniform distribution in the range of $[0, 3]$ m/s, and the direction of each MS is uniformly distributed in the range of $[0, 2\pi]$. To ensure that each MS always locates in the coverage of the cell, once the MS arrives at the boundary of the cell, we let the MS moves along the reverse direction of the current direction. Referring to [33], the channel gain is modeled as $Y / L(l)$, where Y accounts for the loss of shadow fading and follows a lognormal distribution with variance $\sigma_s^2 = 10$ dB. $L(l)$ is the path loss between the transmitter and the receiver, and can be calculated from $10\log_{10}L(l) = 128.1 + 37.6\log_{10}l$, where l (in kilometers) represents the distance between the transmitter and the receiver. The background noise level is set to be −104 dBm.
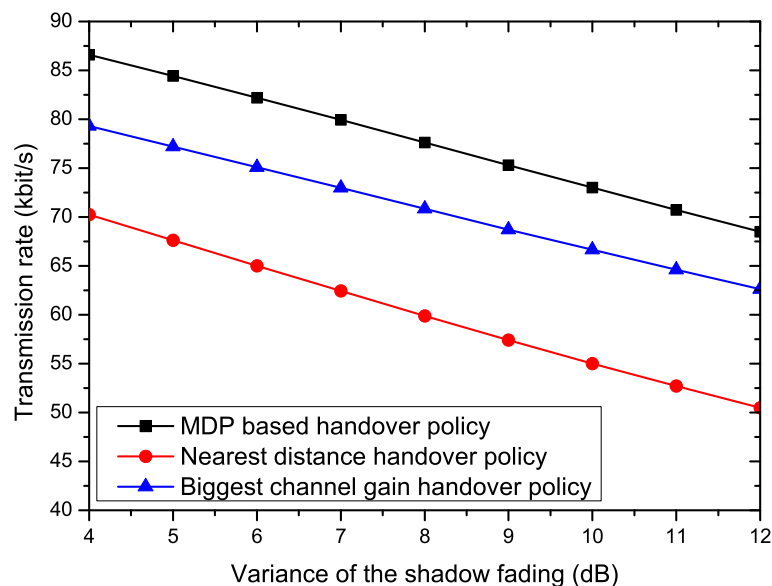

Figure 6 The system transmission rates with different variances of shadow fading.

In the simulation, the set of the modulation and coding schemes consists of 1/2-BPSK, 1/2-QPSK, 2/3-QPSK, 5/6-QPSK, 1/2-16QAM, and 2/3-16QAM, where their thresholds of SNR maintaining the target BER of 10^{-3} can be found in Table 1 [34]. Then, the set of downlink transmission rates can be denoted as

$$\Lambda = \left\{ 0, \frac{1}{2}c, \ c, \ \frac{4}{3}c, \ \frac{3}{2}c, \ \frac{5}{3}c, \ 2c, \ \frac{8}{3}c \right\} \tag{36}$$

where the symbol rate c is set to be 640 ksymbols/s. Moreover, we use a simulation-based method to estimate the state transition probability. For more detailed description of this method, the readers can refer to [32]. The other simulation parameters of the system are summarized in Table 2. In the following, the system transmission rates (i.e., the average transmission rates of every MS) and the expected total reward per connection will be used as the performance metrics.

Figures 5, 6, and 7 illustrate the system transmission rates of the three policies in different scenarios. Specifically, Figure 5 plots the system transmission rates versus the transmit power of the BS, where the transmit power of each RS is set to be 10 mW and the number of MSs in the cellular is set to be 100. It can be observed that, with the increase of the transmit power of the BS, the received SNR of each MS will also increase. Accordingly, a better adaptive modulation and coding scheme will be selected, which results in a greater system transmission rate. Moreover, it can be found that the transmission rates of the proposed MDP-based handover policy are always greater than that of the nearest distance handover policy and the biggest channel gain handover policy, actually, about more than 25 and 10%, respectively. That is to say, the performance of

handover has been improved by employing the proposed MDP-based handover policy in wireless relay systems.

Figure 6 shows the relationship between the system transmission rates and the variance of shadow fading. In this simulation, there are 100 MSs in the system, the transmit power of each RS is fixed to be 10 mW, the transmit power of the BS is set to be 1 W, and the variance of the shadow fading is increasing from 4 dB to 12 dB. It can be observed from Figure 6 that the system transmission rates decrease with the increase of the variance of the shadow fading. The reason is that, when the variance of the shadow fading increases, the channel fading becomes worse, and then the received SNR of each MS will decrease. Accordingly, the system transmission rates will be affected. Moreover, compared with the other two policies, the proposed MDP-based handover policy can consistently achieve the biggest transmission rate, which indicates that the proposed policy is well behaved.

Figure 7 depicts the system transmission rates versus the number of MSs, where the transmit power of the BS is set to be 1 W and the transmit power of each RS is set to be 10 mW. In Figure 7, the transmission rates of the three policies are shown to be decreased with the increase of the number of the MSs. This is due to the fact that, the power resource will be allocated to more MSs with the increase of the number of the MSs. Once again, it should be noted that the proposed MDP-based handover policy always outperforms the other two policies in terms of the system transmission rate, which also proves that the proposed MDP-based handover policy is well behaved.

Figures 8 and 9 illustrate the expected total rewards of the three policies in different scenarios. Specifically, Figure 8 shows the expected total reward versus the
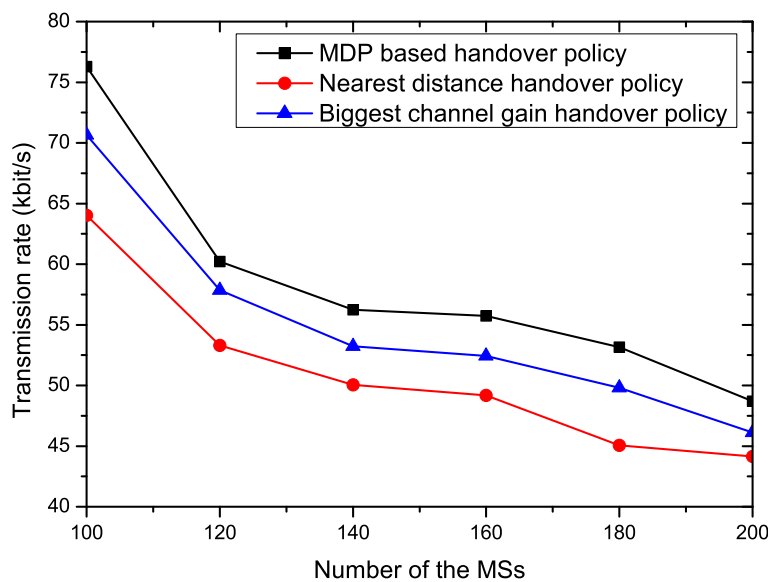

Figure 7 The system transmission rates with different number of MSs.
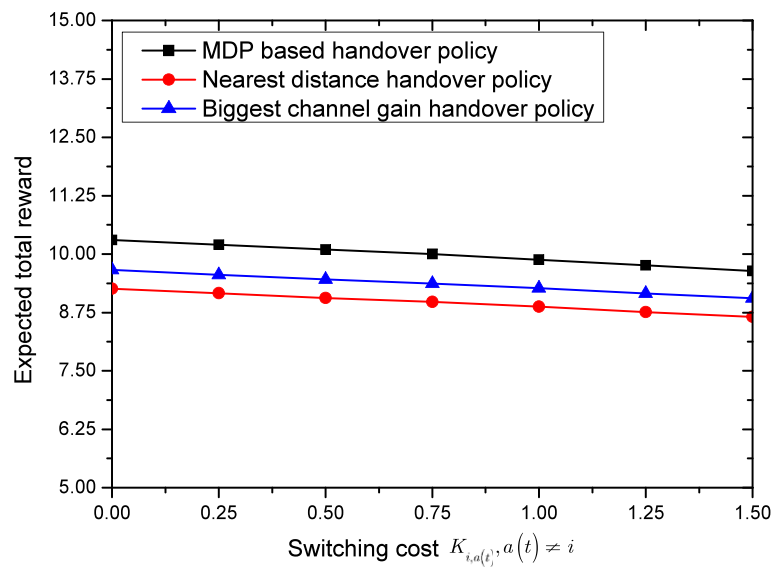
Figure 8 The expected total reward versus the switching cost between every two different serving nodes.

switching cost between every two different serving nodes (i.e., $K_{i,a(t)}, a(t) \neq i$). In Figure 8, with the increase of the switching cost, the expected total rewards of all three policies decrease. This indicates that the switching cost provides flexibility for the system operators. In other words, small values of the switching cost can be set among serving nodes with light signaling load, while large values of the switching cost can temporarily be set for overload serving nodes as a load-balancing technique to decrease the traffic load in the system. Moreover, the proposed MDP-based handover policy always keeps the biggest expected total reward compared to the other two policies, which means the proposed MDP-based handover

policy achieves the best performance in terms of the expected total reward.

Figure 9 shows the expected total reward versus the discount factor λ. It can be seen that the expected total rewards of the three policies increase with the increase of the discount factor. This conclusion can directly be derived from Equation (25) or (27). Moreover, compared to the other two policies, the MDP-based handover policy can consistently provide the highest expected total reward per connection for all values of λ. For example, when the discount factor is 0.9, the MDP-based handover policy can obtain 10 and 4% more total expected total reward than the nearest distance handover policy
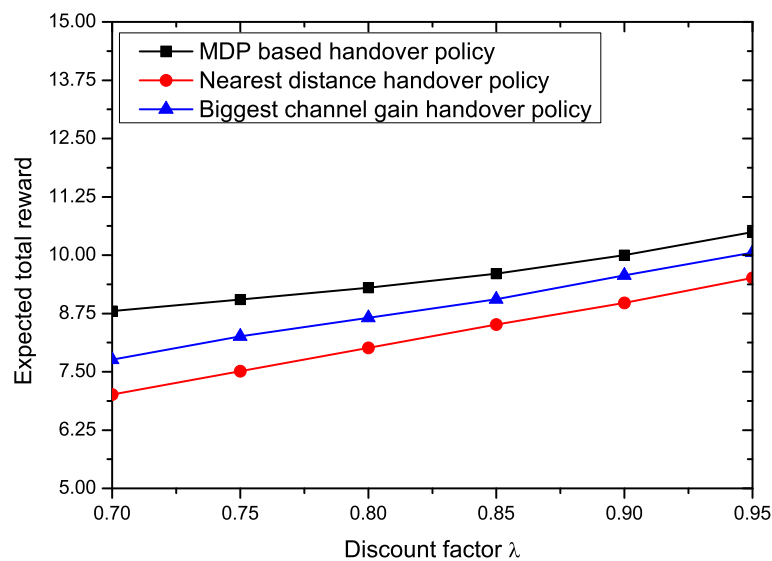

Figure 9 The expected total reward versus the discount factor λ.

and the biggest channel gain handover policy, respectively. Once again, it should be noted that the proposed MDP-based handover policy is well behaved in wireless relay systems.

6 Conclusion

This article has focused on the handover decision problem in wireless relay transmission systems. The architecture of the single-cell relay system is first introduced. Then, the handover decision problem is formulated as an MDP. A profit function is used to evaluate the QoS of the chosen serving node, and a cost function is used to model the signaling cost. Based on the two functions, a reward function is formulated. After that, in order to maximize the expected total reward per connection, a stationary deterministic optimal policy is obtained by employing the VIA. Furthermore, we have shown that the proposed MDP-based handover policy can be generalized to multiple-cell systems. Numerical results show that the MDP-based handover policy proposed in this article is well behaved in solving the handover problem of wireless relay systems.

Abbreviations
BS: Base station; FDMA: Frequency division multiple access; MDP: Markov decision process; MS: Mobile station; OFDMA: Orthogonal frequency division multiple access; QoS: Quality-of-service; RSs: Relay stations; SINR: Signal-to-interference-plus-noise ratio; SNR: Signal-to-noise ratio; VIA: Value iteration algorithm.

Competing interests
The authors declare that they have no competing interests.

Author details
[1]College of Electronic and Information Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China. [2]National Mobile Communications Research Laboratory, Southeast University, Nanjing 210096, China.

References
1. J Wang, J Chen, Performance of wideband CDMA systems with complex spreading and imperfect channel estimation. IEEE J. Sel. Areas Commun. **19**(1), 152–163 (2001)
2. J-B Wang, M Chen, X Wan, C Wei, Ant-colony-optimization-based scheduling algorithm for uplink CDMA nonreal-time data. IEEE Trans. Veh. Technol. **58**(1), 231–241 (2009)
3. H Zhu, J Wang, Chunk-based resource allocation in OFDMA systems—Part: I chunk allocation. IEEE Trans. Commun. **57**(9), 2734–2744 (2009)
4. J-B Wang, M Chen, Optimal uplink packet scheduling for CDMA non-real time data. IEE Electron. Lett. **43**(14), 764–766 (2007)
5. IEEE, Draft Standard for Local and Metropolitan Area Networks—Part 16: Air Interface for Fixed and Mobile Broadband Wireless Access Systems: Multihop Relay Specification. IEEE Standard (2007). 2007. 4312731
6. IEEE, Part 16: Air Interface for Fixed and Mobile Broadband Wireless Access Systems: Multihop Relay Specification, IEEE P802.16j Base Line Document (2007)
7. EC van der Meulen, Three-terminal communication channels. Adv. Appl. Probab. **3**(1), 120–154 (1971)
8. TM Cover, AA El Gamal, Capacity theorems for the relay channel. IEEE Trans. Inf. Theory. **IT-25**(5), 572–584 (1979)
9. X Lei, W Li, Exact closed-form expression for ergodic capacity of amplify-and-forward relaying in channel-noise-assisted cooperative networks with relay selection. IEEE Commun. Lett. **15**(3), 332–333 (2011)
10. Y Ding, J-K Zhang, KM Wong, Ergodic channel capacities for the amplify-and-forward half-duplex cooperative systems. IEEE Trans. Inf. Theory. **55**(2), 713–730 (2009)
11. U-K Kwon, C-H Choi, G-H Im, Full-rate cooperative communications with spatial diversity for half-duplex uplink relay channels. IEEE Trans. Wirel. Commun. **8**(11), 5449–5454 (2009)
12. J-B Wang, Y Jiao, M Chen, J-Y Wang, Z Cao, Multi-hop free space optical communications using serial decode-and-forward relay transmissions. China Commun. **8**(5), 102–110 (2011)
13. Z Ding, KK Leung, DL Goeckel, D Towsley, On the study of network coding with diversity. IEEE Trans. Wirel. Commun. **8**(3), 1247–1259 (2009)
14. J Luo, RS Blum, LJ Cimini, LJ Greenstein, AM Haimovich, Decode-and-forward cooperative diversity with power allocation in wireless networks. IEEE Trans. Wirel. Commun. **6**(3), 793–799 (2007)
15. TC-Y Ng, W Yu, Joint optimization of relay strategies and resource allocations in cooperative cellular networks. IEEE J. Sel. Areas Commun. **25**(2), 328–339 (2007)
16. R Pabst, BH Walke, DC Schultz, P Herhold, H Yanikomeroglu, S Mukherjee, H Viswanathan, M Lott, W Zirwas, M Dohler, H Aghvami, DD Falconer, GP Fettweis, Relay-based deployment concepts for wireless and mobile broadband radio. IEEE Commun. Mag. **42**(9), 80–89 (2004)
17. H Wu, C Qiao, S De, O Tonguz, Integrated cellular and ad hoc relaying systems: iCAR. IEEE J. Sel. Areas Commun. **19**(10), 2105–2115 (2001)
18. S Cho, EW Jang, JM Cioffi, Handover in multihop cellular networks. IEEE Commun. Mag. **47**(7), 64–73 (2009)
19. M Stemm, RH Katz, Vertical handoffs in wireless overlay networks. ACM/Baltzer Mobile Netw. Appl. **3**(4), 335–350 (1998)
20. K Pahlavan, P Krishnamurthy, A Hatami, M Ylianttila, JP Makela, R Pichna, J Vallstron, Handoff in hybrid mobile data networks. IEEE Personal Commun. **7**(2), 34–47 (2000)
21. Y-F Huang, F-B Gao, H-C Hsu, in *Proceedings of the 10th WSEAS International Conference on Signal Processing, Computational Geometry and Artificial Vision*, vol. 1. Performance of an adaptive timer-based handoff scheme for wireless mobile communications, (Stevens Point, Wisconsin, USA, 2010), pp. 46–51
22. H-P Lin, R-T Juang, D-B Lin, Validation of an improved location-based handover algorithm using GSM measurement data. IEEE Trans. Mobile Comput. **4**(5), 530–536 (2005)
23. J Jiang, W Heng, H Zhang, G Wu, H Wang, in *Proceedings of the International Conference on Wireless Communications & Signal Processing*, vol. 1. A novel handover scheme for multi-hop transparent relay network, (Nanjing, China, November 2009), pp. 1–5
24. M Ylianttila, M Pande, J Makela, P Mahonen, in *Proceedings of the IEEE Global Telecommunications Conference*, vol. 6. Optimization scheme for mobile users performing vertical handoffs between IEEE 802.11 and GPRS/EDGE networks, (San Antonio, TX , USA, November 2001), pp. 3439–3443
25. J-B Wang, H Chen, M Chen, J Wang, Cross-layer packet scheduling for downlink multiuser OFDM systems. Sci. China Ser. F - Inf. Sci. **52**(12), 2369–2377 (2009)
26. W Li, J-B Wang, M Chen, Outage probability of dual-hop amplify-and-forward relaying systems over shadowed Nakagami-m fading channel. IEICE Trans. Fund. Electron. Commun. Comput. Sci. **91**(11), 3403–3405 (2008)
27. M Puterman, *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. (Wiley, Hoboken, NJ, 1994)
28. EA Feinberg, A Schwartz, EA Feinberg, *Handbook of Markov decision processes, ch. Markov Decision Processes in Finance and Dynamic Options*. (Kluwer, Netherlands, 2002)
29. F Yu, V Krishnamurthy, VCM Leung, Cross-layer optimal connection admission control for variable bit rate multimedia traffic in packet wireless CDMA networks. IEEE Trans. Signal Process. **54**(2), 542–555 (2006)
30. F Yu, VWS Wong, VCM Leung, Efficient QoS provisioning for adaptive multimedia in mobile communication networks by reinforcement learning. Mobile Netw. Appl. **11**(1), 101–110 (2006)

31. CK Tham, JC Renaud, in *Proceedings of the Second International Conference on Intelligent Sensors, Sensor Networks and Information Processing*, vol. 1. Multi-agent systems on sensor networks: a distributed reinforcement learning pproach, (Melbourne, Australia, December 2005), pp. 423–429

32. E Stevens-Navarro, Y Lin, VWS Wong, An MDP-based vertical handoff decision algorithm for heterogeneous wireless networks. IEEE Trans. Veh. Technol. **57**(2), 1243–1254 (2008)

33. UMTS, *Annex B: Test Environments and Deployment Models*, (1998). UMTS 30.03

34. J Moon, S Hong, in *Proceedings of the IEEE 56th Vehicular Technology Conference*, vol. 4. Adaptive code-rate and modulation for multi-user OFDM system in wireless communications, (Vancouver, Canada, September 2002), pp. 1943–1947