A Constrained MDP-based Vertical Handoff Decision Algorithm for 4G Wireless Networks

Chi Sun, Enrique Stevens-Navarro, and Vincent W.S. Wong
Department of Electrical and Computer Engineering
The University of British Columbia, Vancouver, BC, Canada
E-mail: {chis, enriques, vincentw}@ece.ubc.ca

***Abstract*— The 4th Generation (4G) wireless communication systems aim to provide users with the convenience of seamless roaming among heterogeneous wireless access networks. To achieve this goal, the support of vertical handoff in mobility management is crucial. This paper focuses on the vertical handoff decision algorithm, which determines under what criteria vertical handoff should be performed. The vertical handoff decision problem is formulated as a constrained Markov decision process (CMDP). The objective is to maximize the expected total reward of a connection subject to the expected total access cost constraint. In our model, a benefit function is used to assess the quality of the connection, and a penalty function is used to model signaling and call dropping. The user's velocity and location information are considered when making the handoff decisions. The value iteration and Q-learning algorithms are used to determine the optimal policy. Numerical results show that our proposed vertical handoff decision algorithm outperforms another scheme which does not consider the user's velocity.**

I. Introduction

The goal of the 4th Generation (4G) wireless communication systems is to utilize different access technologies in order to provide multimedia services to users on an "anytime, anywhere" basis. Currently, standardization bodies such as 3rd Generation Partnership Project (3GPP), 3GPP2, and the IEEE 802.21 Media Independent Handover (MIH) working group are working towards this vision. In the 4G communication systems, users will have a variety of wireless networks to choose from in order to send and/or receive their data. A user can either choose to use Universal Mobile Telecommunications System (UMTS) to benefit from a good quality of service (QoS), Worldwide Interoperability for Microwave Access (WiMAX) to achieve a high data rate, or wireless local area network (WLAN) to enjoy a moderate cost. As a result, seamless mobility must be properly managed to achieve the goal of the 4G wireless systems, and *vertical handoff* is a crucial key for supporting seamless mobility.

Vertical handoff support is responsible for service continuity when a connection needs to migrate across heterogeneous wireless access networks. It generally involves three phases [1], [2]: *system discovery*, *vertical handoff decision*, and *vertical handoff execution*. During the system discovery phase, the mobile terminal (MT) receives advertised information from different access networks. These messages may include their access costs and QoS parameters for different services. In the vertical handoff decision phase, the MT determines whether the current connection should keep using the same network or switch to another one. The decision is based on the information it received during the system discovery phase, and the current state conditions (e.g., MT's current location, velocity, battery status). In the vertical handoff execution phase, the connections are seamlessly migrated from the existing network to another. This process involves authentication, authorization, and also the transfer of context information.

Various vertical handoff decision algorithms have been proposed in the literature recently. In [3], the vertical handoff decision is formulated as a fuzzy multiple attribute decision making problem and two methods are proposed: SAW (Simple Additive Weighting) and TOPSIS (Technique for Order Preference by Similarity to Ideal Solution). In [4], an MDP-based vertical handoff decision algorithm is proposed. The problem is formulated as an MDP, but the model does not consider the user's velocity and location information. In [5], a vertical handoff decision algorithm based also on dynamic programming is presented. The model considers the user's location and mobility information but assumes there is no constraint on the user's total budget for each connection. The user's velocity is considered in the vertical handoff decision algorithm proposed in [6]. In [7], a framework is proposed to evaluate different vertical handoff algorithms, in which the MT's mobility is modeled by a Markov chain. In [8], a utility-based network selection strategy is presented. A number of utility functions are examined to capture the tradeoffs between the users' preference and their vertical handoff decisions.

In this paper, we propose a vertical handoff decision algorithm for 4G wireless networks. The problem is formulated as a constrained Markov decision process (CMDP). The objective is to maximize the expected total reward per connection subject to the expected total access cost constraint. The contributions of our work are as follows:

- Our proposed model takes into account the resources available in different networks, and the MT's information (e.g., location, velocity). A benefit function is used to model the bandwidth and delay of the connection. A penalty function is used to model the signaling incurred and the call dropping probability. A cost function is used to capture the access cost of using a specific network.
- We determine the optimal policy for decision making via the use of value iteration and Q-learning algorithms.
- We evaluate the performance of our proposed algorithm under different parameters. Numerical results show that

our proposed vertical handoff decision algorithm outperforms another scheme which does not consider the user's velocity in making the decisions.

The rest of the paper is organized as follows. The system model is presented in Section II. The CMDP formulation and optimality equations are described in Section III. Section IV presents the numerical results and discussions. Conclusions are given in Section V.

II. System Model

In this section, we describe how the vertical handoff decision problem can be formulated as a constrained Markov decision process (CMDP). A CMDP model can be characterized by six elements: *decision epochs, states, actions, transition probabilities, rewards,* and *costs* [9]. At each decision epoch, the MT has to choose an action based on its current state. With this state and action, the MT then evolves to a new state according to a transition probability function. This new state lasts for a period of time until the next decision epoch comes, and then the MT makes a new decision again. For any action that the MT chooses at each state, there is a reward and a cost associated with it. The goal of each MT is to maximize the expected total reward it can obtain during the connection lifetime, subject to the expected total access cost constraint.

A. States, Actions and Transition Probabilities

We represent the decision epochs by $T = \{1, 2, \ldots, N\}$, where the random number N indicates the time that the connection terminates. We denote the state space of the MT by S, and we only consider finite number of states that an MT can possibly be in. The state of the MT contains information such as the current network that the MT connects to, the available bandwidth and delay that all the networks offer, and the velocity and location information of the MT. Specifically, the state space can be expressed as follows:

$$S = M \times B^1 \times D^1 \times \cdots \times B^{|M|} \times D^{|M|} \times V \times L,$$

where $\times$ denotes the Cartesian product, M represents the set of available network IDs that the MT can connect to. B^m and D^m, where $m \in M$, denote the set of available bandwidth and delay of network m, respectively. V denotes the set of possible velocity values of the MT, and L denotes the set of location type (LT) that the MT can possibly reside in.

Since a finite countable state space is being considered in this paper, the bandwidth and delay can be quantized into multiple of unit bandwidth and unit delay, respectively [9]. Specifically, for network $m \in M$, the set of available bandwidth $B^m = \{1, 2, \ldots, b^m_{max}\}$, where b^m_{max} denotes the maximum bandwidth available to a connection from network m. For example, the unit bandwidth of WLAN and the UMTS network can be 500 *kbps* and 16 *kbps*, respectively.

Similarly, for network $m \in M$, the set of available delay $D^m = \{1, 2, \ldots, d^m_{max}\}$, where d^m_{max} denotes the maximum delay provided to a connection by network m. For example, the unit delay of WLAN and the UMTS network can be 50 *ms* and 20 *ms*, respectively.

The velocity of the MT is also quantized as multiple of unit velocity. The set of possible velocity values is

$$V = \{0, 1, 2, \ldots, v_{max}\},$$

where v_{max} denotes the maximum velocity that an MT can travel at. For example, the unit of velocity can be 10 *km/h*.

For the set of location type (LT) that the MT can possibly reside in, we have:

$$L = \{1, 2, \ldots, l_{max}\},$$

where l_{max} denotes the total number of different LTs in the area of interest. LTs are differentiated by the number of networks they are covered by.

Let vector $\mathbf{s} = [i, b_1, d_1, \ldots, b_{|M|}, d_{|M|}, v, l]$ denote the current state of the MT, where i denotes the current network used by the connection, b_m and d_m denote the current bandwidth and delay of network m, respectively, v denotes the current velocity of the MT, and l denotes the current LT that the MT resides in. At each decision epoch, based on the current state $\mathbf{s}$, the MT chooses an action $a \in A_{\mathbf{s}}$, where the action set $A_{\mathbf{s}} \subset M$ consists of the IDs of the network that the MT can potentially switch to. If the chosen action is a, the probability that the next state $\mathbf{s}' = [j, b'_1, d'_1, \ldots, b'_{|M|}, d'_{|M|}, v', l']$ is:

$$P[\mathbf{s}'|\mathbf{s}, a] = \begin{cases} P[v'|v]P[l'|l]\prod_{m\in M} P[b'_m, d'_m|b_m, d_m], & j = a, \\ 0, & j \neq a, \end{cases} \quad (1)$$

where $P[v'|v]$ is the transition probability of the MT's velocity, $P[l'|l]$ is the transition probability of the MT's LT, and $P[b'_m, d'_m|b_m, d_m]$ is the joint transition probability of the bandwidth and delay of network m.

The transition probability of the MT's velocity is obtained based on the Gauss-Markov mobility model from [10]. In this model, an MT's velocity is assumed to be correlated in time and can be modeled by a discrete Gauss-Markov random process. The following recursive realization is used to calculate the transition probability of the MT's velocity:

$$v' = \alpha v + (1 - \alpha)\mu + \sigma\sqrt{1 - \alpha^2}\phi, \quad (2)$$

where v is the MT's velocity at the current decision epoch, v' is the MT's velocity at the next decision epoch, α is the memory level (i.e., $0 \leq \alpha \leq 1$), μ and σ are the mean and standard deviation of v, respectively, and ϕ is an uncorrelated Gaussian process with zero mean and unit variance (i.e., $\phi \sim N(0, 1)$) which is independent of v. By varying v and counting the number of different outcomes of v' according to (2), the MT's velocity transition probability matrix (i.e., $P[v'|v]$) can be obtained in a simulation-based manner.

For the transition probability of the MT's LT, we assume that an access network which has a smaller coverage area (e.g., WLAN) always lies within another network that has a larger coverage area (e.g., WiMAX). Although this assumption might not hold for the cases when M is large, it is still reasonable if the number of different networks does not exceed three, which is a typical case in today's wireless communication systems.

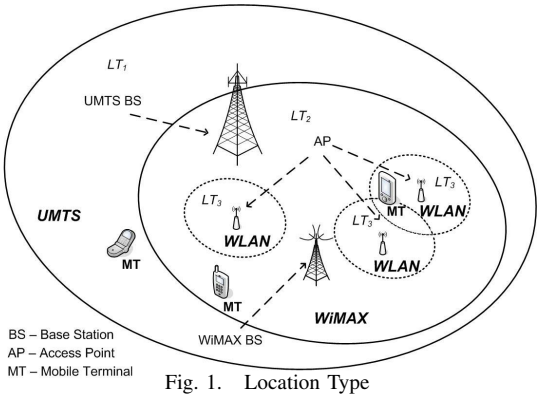

Fig. 1. Location Type

We define LT_l, where $l \in L$, to be the area covered by networks $\{1, \ldots, l\}$ but not covered by networks $\{l+1, \ldots, l_{max}\}$. For example, in Fig. 1, l_{max} is three since the number of different LTs in the system is equal to three. We assign the IDs of UMTS, WiMAX, and WLAN to be 1, 2, and 3, respectively. LT_1 is the area covered only by the UMTS network, LT_2 is the area covered by UMTS and WiMAX, but not WLAN, and LT_3 is the area covered by all three networks (i.e., UMTS, WiMAX, and WLAN). Under this assumption, the number of different LTs (i.e., l_{max}) is essentially equal to the number of different networks (i.e., $|M|$) in the system.

Let A_{LT_l} denote the total area of LT_l and ρ_l denote the user density of LT_l. The effective area of LT_l is:

$$A^E_{LT_l} = A_{LT_l}\, \rho_l. \tag{3}$$

In real world, the user density in different networks (e.g., WLAN and the UMTS network) are not the same [11], [12], so the density index of each LT is put into consideration to achieve a more realistic model.

We assume that an MT currently at LT_l can only move to its neighboring LTs (i.e., either LT_{l+1} or LT_{l-1}) or stay at LT_l at the next decision epoch. This is because the duration of each decision epoch is too short for the MT to traverse more than one LT areas. Thus, the probability that an MT's next LT is $LT_{l'}$ given its current LT is LT_l is assumed to be proportional to the effective area of $LT_{l'}$. Specifically, the transition probability of an MT's LT is defined as follows:

$$P[l'|l] = \begin{cases} \frac{A^E_{LT_{l'}}}{\sum_{\xi=l,l+1} A^E_{LT_\xi}}, & \text{if } l = 1, \\ \frac{A^E_{LT_{l'}}}{\sum_{\xi=l-1,l,l+1} A^E_{LT_\xi}}, & \text{if } l = 2, \ldots, l_{max}-1, \\ \frac{A^E_{LT_{l'}}}{\sum_{\xi=l-1,l} A^E_{LT_\xi}}, & \text{if } l = l_{max}. \end{cases} \tag{4}$$

Note that the LT where an MT resides in determines its action set. If an MT is at LT_l, its action set $A_{\mathbf{s}}$ only contains the entries from 1 to l.

B. Rewards

When an MT chooses an action a in state $\mathbf{s}$, it receives an immediate reward $r(\mathbf{s}, a)$. The reward function depends on the benefit function and the penalty function, which are explained below.

For the benefit function of the MT, two aspects are considered: bandwidth and delay. Let the *bandwidth benefit function* represent the benefit that an MT can gain (in terms of bandwidth) by selecting action a in state $\mathbf{s}$:

$$f_b(\mathbf{s}, a) = \begin{cases} 1, & \text{if } b_i = \max_{k \in M}\{b_k\}, \quad a = i, \\ 0, & \text{if } b_i = \max_{k \in M}\{b_k\}, \quad a \neq i, \\ \frac{b_a - b_i}{\max_{k \in M}\{b_k - b_i\}}, & \text{if } b_i \neq \max_{k \in M}\{b_k\}, \quad b_a > b_i, \\ 0, & \text{if } b_i \neq \max_{k \in M}\{b_k\}, \quad b_a \leq b_i. \end{cases}$$

The benefit is being assessed as follows. Given that the MT is currently connecting to network i. If network i is the one which offers the highest bandwidth among others, the strategy is to keep using network i. However, if the MT is not using the network which has the highest bandwidth, the benefit that it can obtain is represented by a fraction, in which the numerator is the MT's actual increase of bandwidth by choosing action a in state $\mathbf{s}$, and the denominator is the MT's maximum possible increase of bandwidth.

Similarly, a *delay benefit function* is used to represent the benefit that an MT can gain (in terms of delay) by choosing action a in state $\mathbf{s}$:

$$f_d(\mathbf{s}, a) = \begin{cases} 1, & \text{if } d_i = \min_{k \in M}\{d_k\}, \quad a = i, \\ 0, & \text{if } d_i = \min_{k \in M}\{d_k\}, \quad a \neq i, \\ \frac{d_i - d_a}{\max_{k \in M}\{d_i - d_k\}}, & \text{if } d_i \neq \min_{k \in M}\{d_k\}, \quad d_a < d_i, \\ 0, & \text{if } d_i \neq \min_{k \in M}\{d_k\}, \quad d_a \geq d_i. \end{cases}$$

As a result, the total *benefit function* is given by:

$$f(\mathbf{s}, a) = \omega f_b(\mathbf{s}, a) + (1 - \omega) f_d(\mathbf{s}, a), \tag{5}$$

where ω is the importance weight given to the bandwidth aspect with $0 \leq \omega \leq 1$.

We consider two factors for the penalty of the MT. First, the *switching cost penalty function* is represented by:

$$g(\mathbf{s}, a) = \begin{cases} K_{i,a}, & \text{if } i \neq a, \\ 0, & \text{if } i = a, \end{cases} \tag{6}$$

where $K_{i,a}$ is the switching cost from network i to network a. This penalty function captures the processing and signaling load incurred when the connection is migrated from one network to another.

Second, we define the *call dropping penalty function* as:

$$q(\mathbf{s}, a) = \begin{cases} 0, & \text{if } i = a, \\ 0, & \text{if } i \neq a, \quad 0 < v \leq V_{min}, \\ \frac{v - V_{min}}{V_{max} - V_{min}}, & \text{if } i \neq a, \quad V_{min} < v < V_{max}, \\ 1, & \text{if } i \neq a, \quad v \geq V_{max}, \end{cases}$$

where V_{max} and V_{min} denote the maximum and minimum velocity thresholds, respectively. When MT moves faster, the probability that the connection will be dropped during vertical handoff process increases.

The total *penalty function* of an MT is given by:

$$h(\mathbf{s}, a) = g(\mathbf{s}, a) + rq(\mathbf{s}, a), \tag{7}$$

where $r \in [0, 1]$ is the MT's risky index. This factor accounts for user's preferences. Some users allow vertical handoff in order to obtain better QoS although there is a risk that the connection may be dropped during handoff, whereas some others may refrain from switching.

Finally, between two successive vertical handoff decision epochs, the *reward function* is defined as:

$$r(\mathbf{s}, a) = f(\mathbf{s}, a) - h(\mathbf{s}, a). \tag{8}$$

C. Costs

For each period of time that the MT uses network n, it will incur the following access cost (in monetary units per second):

$$c(\mathbf{s}, a) = \begin{cases} \psi_n, & \text{if } a = n, \\ 0, & \text{otherwise}, \end{cases} \tag{9}$$

and for each network n where $n \in M$, we have:

$$\psi_n = b_n \, C_n, \tag{10}$$

where b_n is the available bandwidth in *bps* and C_n is the access cost of network n in monetary units per bit. The user has a budget such that it is willing to spend up to C_{max} monetary units per connection.

III. CMDP Formulation and Optimality Equations

In this section, we present the problem formulation and describe how to obtain the optimal policy. First, some concepts need to be clarified. The random variable N, which denotes the *connection termination time*, is assumed to be geometrically distributed with mean $1/(1-\lambda)$, where λ can also be interpreted as the *discount factor* of the model ($0 \leq \lambda < 1$).

A *decision rule* is a regulation specifying the action selection for each state at a particular decision epoch. It can be expressed as $\delta_t : S \rightarrow A$. A *policy* $\pi = (\delta_1, \delta_2, \ldots, \delta_N)$ is a sequence of decision rules to be used at all N decision epochs.

Let $v^{\pi}(\mathbf{s})$ denote the *expected discounted total reward* between the first decision epoch and the connection termination, given that policy π is used with initial state $\mathbf{s}$. We can state the CMDP optimization problem as:

$$\begin{aligned} \text{maximize} \quad & v^{\pi}(\mathbf{s}) = E_{\mathbf{s}}^{\pi} \left\{ \sum_{t=1}^{\infty} \lambda^{t-1} \, r(\mathbf{s}_t, a_t) \right\}, \\ \text{subject to} \quad & C^{\pi}(\mathbf{s}) = E_{\mathbf{s}}^{\pi} \left\{ \sum_{t=1}^{\infty} \lambda^{t-1} \, c(\mathbf{s}_t, a_t) \right\} \leq C_{max}, \end{aligned} \tag{11}$$

where $E_{\mathbf{s}}^{\pi}$ denotes the expectation with respect to policy π and initial state $\mathbf{s}$, and $C^{\pi}(\mathbf{s})$ denotes the *expected discounted total access cost* calculated using policy π and initial state $\mathbf{s}$.

Since the optimization problem is to maximize the expected discounted total reward, we define a policy π^* to be *optimal* in Π if $v^{\pi^*}(\mathbf{s}) \geq v^{\pi}(\mathbf{s})$ 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 $\pi = (\delta, \delta, \ldots, \delta)$, and for convenience we denote π simply by δ. A policy is said to be *deterministic* if it chooses an action with certainty at each decision epoch. We refer to stationary deterministic policies as *pure* policies [13].

To solve (11), we can use the *Lagrangian approach* [9], [14] to reduce it into an equivalent unconstrained MDP problem. By including the Lagrange multiplier β with $\beta > 0$, we have:

$$r(\mathbf{s}, a; \beta) = r(\mathbf{s}, a) - \beta c(\mathbf{s}, a). \tag{12}$$

Then, the *optimality equations* are given by:

$$v_{\beta}(\mathbf{s}) = \max_{a \in A_{\mathbf{s}}} \left\{ r(\mathbf{s}, a; \beta) + \sum_{\mathbf{s}' \in S} \lambda \; P[\mathbf{s}' | \mathbf{s}, a] \; v_{\beta}(\mathbf{s}') \right\}, \tag{13}$$

which can be solved by using the Value Iteration Algorithm (VIA) [13] with a fixed value of β. The solutions of (13) correspond to the maximum expected discounted total reward $v_{\beta}(\mathbf{s})$ and the pure policy δ_{β}. Note this pure policy δ_{β} specifies the network to choose in each state $\mathbf{s}$, such that the expected discounted total reward is maximized.

The Q-learning algorithm proposed in [14] is used to determine the proper β (i.e., β^*) for a feasible C_{max}. Specifically, the iteration algorithm is described by the following equation:

$$\beta_{k+1} = \beta_k + \frac{1}{k}(C^{\delta_{\beta}} - C_{max}) \tag{14}$$

where k is the iteration number.

Once β^* has been obtained, we follow the procedures in [14] to find the optimal policy for the CMDP problem. As discussed in [15], the optimal policy for a CMDP with single constraint is a mixed policy of two pure policies. First, we perturb β^* by some $\Delta\beta$ to get $\beta^- = \beta^* - \Delta\beta$ and $\beta^+ = \beta^* + \Delta\beta$. Then, we calculate the pure policies δ^- and δ^+ (using β^- and β^-, respectively) and their corresponding expected discounted total access costs $C^- = C^{\delta^-}$ and $C^+ = C^{\delta^+}$. Next, we define a parameter q such that $qC^- + (1-q)C^+ = C_{max}$. The *optimal policy* δ^* of the CMDP is a *randomized* mixture of two policies (i.e., δ^- and δ^+), such that at each decision epoch, the first policy is chosen with probability q and the second one is chosen with probability $1-q$. In other words, the optimal policy can be described as follows:

$$\delta^* = q\delta^- + (1-q)\delta^+ \tag{15}$$

IV. Numerical Results and Discussions

We compare the performance between our proposed CMDP-based vertical handoff decision algorithm with another scheme, which is also based on the CMDP but does not consider the impact on velocity in making the decisions (this scheme is denoted by CMDP-w/o-velocity). The performance metric is the *expected total reward per connection*. The application considered is constant bit rate (CBR) voice traffic using the user datagram protocol (UDP) as the transport protocol.

We consider the scenario that there are two networks in the system: network 1 is the cellular network and network 2 is WLAN. The average duration between two successive decision epochs is 15 $secs$. For both networks, the unit of bandwidth and delay are equal to 16 $kbps$ and 60 ms, respectively. The

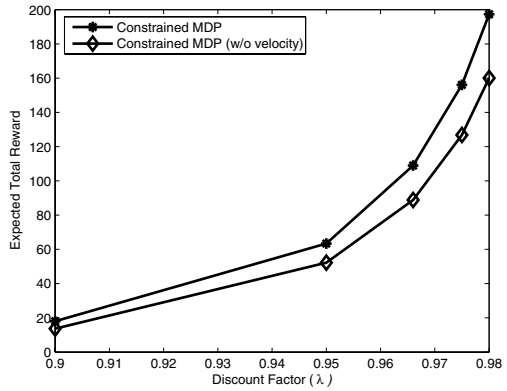

Fig. 2. Expected total reward under different discount factor (λ).

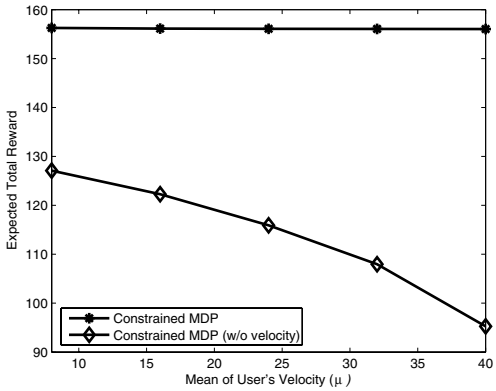

Fig. 3. Expected total reward under different mean of user's velocity (μ).

maximum available bandwidth and delay in network 1 (i.e., b^1_{max} and d^1_{max}) and network 2 (i.e., b^2_{max} and d^2_{max}) are 5 units, 4 units, 15 units, and 4 units, respectively. The unit of the MT's velocity is 8 km/h, and the maximum possible velocity of the MT is 5 units, with the lower and upper thresholds (i.e., V_{min} and V_{max}) equal to 1 unit and 5 units, respectively. For the Gauss-Markov model, the memory level α is 0.5, the standard deviation of the MT's velocity σ is 0.1 unit, and the mean of the MT's velocity μ is equal to 1 unit. The area of LT_1 and LT_2 are assumed to be 75% and 25% of the total area [16], respectively. The ratio between the user densities ρ_1:ρ_2 = 1:8. The switching cost $K_{1,2} = K_{2,1} = 0.5$. The importance weight ω is 0.25, as CBR traffic is more sensitive to delay. The risky index r of the MT is 0.5. The access cost of networks 1 and 2 are 3 and 1 monetary units per bit, respectively.

For the cellular network, the values of bandwidth and delay are assumed to be guaranteed for the duration of the connection (i.e., $P[b_1, d_1|b_1, d_1] = 1$). For WLAN, we estimate such probabilities in a simulation-based manner. In ns-2 simulator [17], a typical IEEE 802.11b WLAN is simulated in which the users arrive and depart from the network with an average Poisson rate of 0.2 users per second. The resulting available bandwidth and delay are rounded according to the predefined units, and the counting of transitions among states is performed to estimate the state transition probability of WLAN (i.e., $P[b'_2, d'_2|b_2, d_2]$).

The probability q that determines the randomized optimal policy in (15) is calculated for different discount factors (i.e., different average connection durations). Specifically, for λ equals to [0.9, 0.95, 0.966, 0.975, 0.98], the corresponding probabilities q are [0.18, 0.54, 0.66, 0.57, 0.60]. Moreover, the user's budget on the expected total access cost is also predefined for different discount factors. Specifically, for λ equals to [0.9, 0.95, 0.966, 0.975, 0.98], the corresponding constraints C_{max} are [92, 194, 294, 388, 466].

The expected total reward of users under different discount factors are shown in Fig. 2. The expected total reward increases as λ becomes larger. This is because the larger λ is, the longer the average duration of the connection becomes. With the same constraint on the expected total access cost, the CMDP algorithm achieves a higher expected total reward

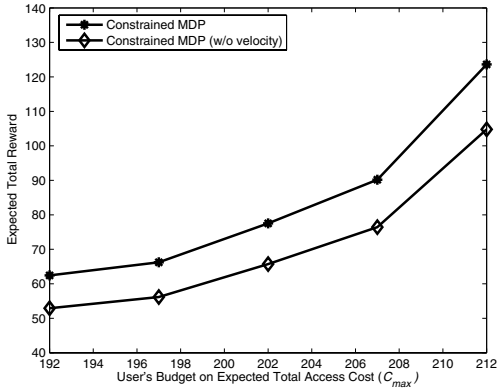

Fig. 4. Expected total reward under different user's budget on expected total access cost (C_{max}).

than the CMDP-w/o-velocity scheme does. For example, when λ equals to 0.975 (i.e., the average duration of connection is 600 *secs*), for which the predefined constraint is 388 monetary units, the CMDP algorithm achieves 23% higher expected total reward than the CMDP-w/o-velocity algorithm does. The reason is that when an MT does not consider its velocity, the connection might be dropped during the handoff process and needs to be re-established. The associated QoS degradation and extra signaling and processing costs decrease the actual reward it will gain by performing the handoff.

Fig. 3 shows the expected total reward of a user versus the mean of its velocity. As the user moves faster, the expected total reward that the CMDP algorithm achieves remains unchanged. This is because the CMDP algorithm effectively avoids dropped calls by taking the user's velocity into consideration. For example, handoffs are only performed when the user's velocity is not likely to cause a dropped call. For the CMDP-w/o-velocity algorithm, the expected total reward decreases as the user's velocity increases. The reason is that as the user becomes faster, the decrease in the actual reward (e.g., QoS degradation and extra signaling and processing costs) associated with the issue that the model does not consider the effect of user's velocity becomes more significant.

The expected total reward a user can obtain versus its budget on the expected total access cost is shown in Fig. 4. As the user's budget increases, the expected total reward becomes

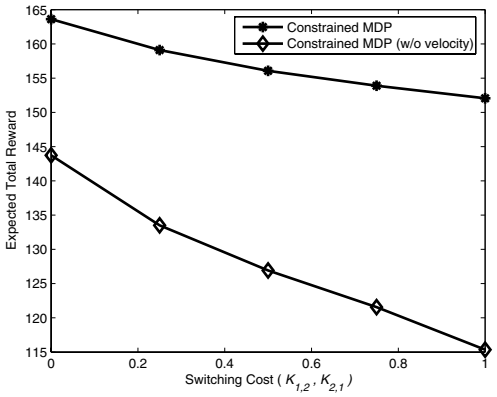

Fig. 5. Expected total reward under different switching cost ($K_{1,2}, K_{2,1}$).

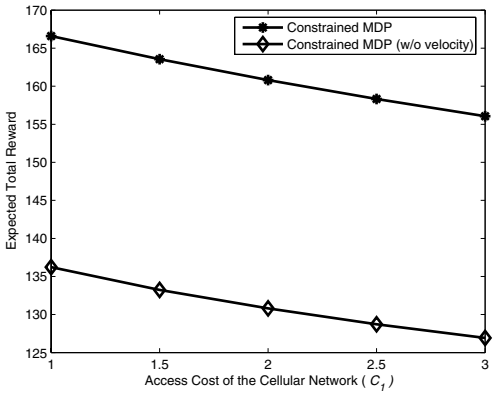

Fig. 6. Expected total reward under different access cost of the cellular network (C_1).

larger. The reason is that the more money that a user can spend on a connection, the more reward it will obtain. For the same budget, the CMDP algorithm always achieves a higher reward than the CMDP-w/o-velocity scheme does. The reason is the CMDP algorithm can fully utilize the user's budget and avoid dropped calls to achieve the optimal reward, while the total reward obtained by the CMDP-w/o-velocity scheme is reduced because of the dropped connections.

Fig. 5 shows the expected total reward under different switching costs. When $K_{1,2}$ and $K_{2,1}$ increase, the expected total reward of both schemes decrease. The expected total reward of the CMDP algorithm decreases slower than the CMDP-w/o-velocity algorithm does. This is because as the switching costs increase, the decrease on the actual reward achieved by the CMDP-w/o-velocity scheme is also larger. Since for the same number of dropped calls, the extra signaling and processing costs increase as $K_{1,2}$ and $K_{2,1}$ increase.

Fig. 6 shows the expected total reward of a user versus the access cost of the cellular network. As C_1 increases (while C_2 is fixed), the expected total reward becomes smaller for both algorithms. The reason is that in order to take advantage of the cellular network, users need to pay more as the price of the cellular network increases. This can also be viewed as the user's budget becomes smaller. Thus, the expected total reward of the user decreases. For the same constraint on the expected total access cost, the CMDP scheme achieves a better expected total reward than the CMDP-w/o-velocity scheme does.

V. Conclusions

In this paper, we propose a vertical handoff decision algorithm for 4G wireless networks. Our work considers the connection duration, QoS parameters, mobility and location information, network access cost, and the signaling load incurred on the network for the vertical handoff decision. The algorithm is based on CMDP formulation with the objective of maximizing the expected total reward of a connection. The constraint of the problem is on the user's budget for the connection. A stationary randomized policy is obtained when the connection termination time is geometrically distributed. Numerical results show that our CMDP-based algorithm outperforms another scheme which does not consider the user's velocity in making the decisions.

Acknowledgment

This work was supported by Bell Canada and the Natural Sciences and Engineering Research Council of Canada.

References

[1] J. McNair and F. Zhu, "Vertical Handoffs in Fourth-generation Multinetwork Environments," *IEEE Wireless Communications*, vol. 11, no. 3, pp. 8–15, June 2004.

[2] W. Chen, J. Liu, and H. Huang, "An Adaptive Scheme for Vertical Handoff if Wireless Overlay Networks," in *Proc. of ICPAD'04*, Newport Beach, CA, July 2004.

[3] W. Zhang, "Handover Decision Using Fuzzy MADM in Heterogeneous Networks," in *Proc. of IEEE WCNC'04*, Atlanta, GA, March 2004.

[4] E. Stevens-Navarro, Y. Lin, and V. W. S. Wong, "An MDP-based Vertical Handoff Decision Algorithm for Heterogeneous Wireless Networks," *IEEE Trans. on Vehicular Technology*, in press, 2008.

[5] J. Zhang, H. C. Chan, and V. Leung, "A Location-Based Vertical Handoff Decision Algorithm for Heterogeneous Mobile Networks," in *Proc. of IEEE Globecom'06*, San Francisco, CA, November 2006.

[6] Q. Guo, J. Zhu, and X. Xu, "An Adaptive Multi-criteria Vertical Handoff Decision Algorithm for Radio Heterogeneous Networks," in *Proc. of IEEE ICC'05*, Seoul, Korea, May 2005.

[7] A. Zahran and B. Liang, "Performance Evaluation Framework for Vertical Handoff Algorithms in Heterogeneous Networks," in *Proc. of IEEE ICC'05*, Seoul, Korea, May 2005.

[8] O. Ormond, J. Murphy, and G. Muntean, "Utility-based Intelligent Network Selection in Beyond 3G Systems," in *Proc. of IEEE ICC'06*, Istanbul, Turkey, June 2006.

[9] E. Altman, *Constrained Markov Decision Processes*. Chapman and Hall, 1999.

[10] B. Liang and Z. Haas, "Predictive Distance-Based Mobility Management for Multidimensional PCS Networks," *IEEE/ACM Trans. on Networking*, vol. 11, no. 5, pp. 718–732, October 2003.

[11] S. Tang and W. Li, "Performance Analysis of the 3G Network with Complementary WLANs," in *Proc. of IEEE Globecom'05*, St. Louis, MO, November 2005.

[12] A. Doufexi, E. Tameh, A. Nix, S. Armour, and A. Molina, "Hotspot Wireless LANs to Enhance the Performance of 3G and Beyond Cellular Networks," *IEEE Commun. Mag.*, vol. 41, no. 7, pp. 58–65, July 2003.

[13] M. Puterman, *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. John Wiley and Sons, 1994.

[14] V. Djonin and V. Krishnamurthy, "Q-Learning Algorithms for Constrained Markov Decision Process with Randomized Monotone Policies: Application to MIMO Transmission Control," *IEEE Trans. on Signal Processing*, vol. 55, no. 5, pp. 2170–2181, May 2007.

[15] F. Beutler and K. Ross, "Optimal Policies for Controlled Markov Chains with a Constraint," *J. Math. Anal. Appl.*, vol. 112, pp. 236–252, 1985.

[16] H. Liu, H. Bhaskaran, D. Raychaudhuri, and S. Verma, "Capacity Analysis of A Cellular Data System with 3G/WLAN Interworking," in *Proc. of IEEE VTC'03*, Orlando, FL, October 2003.

[17] The Network Simulator - ns-2, http://www.isi.edu/nsnam/ns.