Q-Learning-Based Multirate Transmission Control Scheme for RRM in Multimedia WCDMA Systems

Yih-Shen Chen, *Student Member, IEEE*, Chung-Ju Chang, *Senior Member, IEEE*, and Fang-Chin Ren, *Member, IEEE*

***Abstract*—In this paper, a Q-learning-based multirate transmission control (Q-MRTC) scheme for radio resource management in multimedia wide-band code-division multiple access (WCDMA) communication systems is proposed. The multirate transmission control problem is modeled as a Markov decision process where the transmission cost is defined in terms of the quality-of-service (QoS) parameters for enhancing spectrum utilization subject to QoS constraint. We adopt a real-time reinforcement learning algorithm, called Q-learning, to accurately estimate the transmission cost for the MRTC. In the meantime, we successfully employ the feature extraction method and radial basis function network (RBFN) for the Q-function that maps the original state space into a feature vector that represents the resultant interference profile. The state space and memory-storage requirement are then reduced and the convergence property of the Q-learning algorithm is improved. Simulation results show that the Q-MRTC for a multimedia WCDMA system can achieve higher system throughput by an amount of 80% and better users' satisfaction than the interference-based MRTC scheme, while the QoS requirements are guaranteed. Also, compared to the table-lookup method, the storage requirement is reduced by 41%.**

***Index Terms*—Code-division multiple access (CDMA) communication system, multirate transmission, Q-learning, radio resource management.**

I. Introduction

WIDE-BAND code-division multiple access (WCDMA) is one of the promising radio-access technologies for IMT-2000. The objective of a multimedia WCDMA system is to provide users with radio link access to services comparable to those currently offered by fixed networks, resulting in a seamless convergence of both fixed and mobile services. Different types of services, such as voice, data, image, and compressed video, are integrated in the multimedia WCDMA system. Therefore, an adequate radio resource management (RRM) is required to enhance spectrum utilization while meeting the quality-of-service (QoS) requirements of heterogeneous services. In this paper, the multirate transmission control (MRTC) scheme for RRM in the WCDMA systems is studied.

The MRTC in the multimedia WCDMA system is to assign *power* and *processing gain* to service requests to maximize spectrum utilization and to fulfill QoS requirements and users' satisfaction. In [1], Choi and Shin proposed an uplink CDMA system architecture to provide diverse QoS guarantees for heterogeneous traffic: real-time (class I) traffic and nonreal-time (class II) traffic. They theoretically derived the admission region of real-time connections, transmission power allocation, and the optimum target signal-to-interference ratio (SIR) of nonreal-time traffic so as to maximize the system throughput and satisfy the predefined QoS of heterogeneous traffic.

There is no absolute number of maximum available channels in the WCDMA system because a WCDMA system is interference limited; its capacity is affected by multiple access interference (MAI), which is a function of the number of active users, users' location, channel impairments, and heterogeneous QoS requirements. Much research on CDMA capacity estimation is based on MAI and other considerations [2]–[5]. In [2], a single-service CDMA network with respect to MAI caused by users in the same and adjacent cells was studied. In [3], Huang and Bhargava investigated the uplink performance of a slotted direct-sequence CDMA (DS-CDMA) system providing voice and data services. A lognormal-distributed MAI model was proposed to estimate the remaining capacity in the CDMA system, where its mean and variance were given by a function of the number of users and the mean and variance of each service type. However, in multimedia WCDMA systems, the measured MAI value may not be stationary and it may also be affected by user locations and service profiles. Hämäläinen and Valkealahti [4] proposed an MAI estimation method to facilitate load control, admission control, and packet scheduling. Kim and Honig [5] studied resource allocation for multiple classes of traffic in a single-cell DS-CDMA system. A joint optimization was investigated over the power and processing gain of the multiple classes to determine flexible resource allocation for each user subject to QoS constraints.

Shin *et al.* proposed an interference-based channel assignment scheme for DS-CDMA cellular systems [6]. A channel is assigned if the interference is less than an allowed level, which is determined by the network, subject to the QoS constraints. Instead of a fixed system capacity, this interference-based scheme can adaptively assign a channel according to the actual system capacity dependent upon interference such that the system utilization and grade of service can be improved. The interference-based scheme was further extended to call admission control (CAC) in multimedia CDMA cellular systems [7], [8]. Dim-

Manuscript received April 18, 2002; revised August 18, 2003 and October 17, 2003. This work was supported by the National Science Council, Taiwan, under Contract NSC 90-2219-E-009-006; Chung Hwa Telecommunication Laboratory, Hsinchu, Taiwan, under Contract TL-90-1101; and Lee & MTI Center for Networking Research, National Chiao Tung University, Hsinchu, Taiwan.

Y.-S. Chen and C.-J. Chang are with the National Chiao Tung University, Hsinchu, Taiwan (e-mail: atlas@bn3.cm.nctu.edu.tw; cjchang@cc.nctu.edu.tw).

F.-C. Ren is with the Industrial Technology Research Institute, Hsinchu, Taiwan. (e-mail: Frank_Ren@itri.org.tw).

Digital Object Identifier 10.1109/TVT.2003.822330

itriou and Tafazolli [7] developed a mathematical model to determine the outage limits of a multiple-service CDMA system and to achieve the maximum aggregated capacity for different system parameters. Phan-Van and Luong [8] proposed a soft-decision CAC (SCAC) scheme, where the upper and lower bounds of the interference-limited WCDMA system capacity are derived. In the SCAC, the new call request obtains an admission grant according to a predefined probability function when the system operates between the upper and the lower bounds of the system capacity.

Maximizing spectrum utilization (revenue) while meeting QoS constraints suggests a constrained Markov decision process (MDP) [13] or semi-Markov decision process (SMDP) [9], [10]. These methodologies have been successfully applied to solve many network control problems; however, they require extremely large state space to model these problems exactly. Consequently, the numerical computation is intractable due to the curse of dimensionality. Also, *a priori* knowledge of state-transition probabilities is required. Alternatively, many researchers turned to use the reinforcement learning (RL) algorithms to solve the large state space problems [11]–[14]. The most obvious advantage of the RL algorithm is that it could approach an optimal solution from the online operation if the RL algorithm is converged.

In this paper, we propose a Q-learning-based MRTC (Q-MRTC) scheme for RRM in the multimedia WCDMA systems to maximize the system utilization and to fulfill the users' satisfaction, subject to QoS requirements of packet error probability and packet transmission delay. For the interference-limited system, the system interference profile is chosen as system state and the multirate transmission control is modeled as a total expected discounted problem. Also, an evaluation function is defined to appraise the cumulative cost of the consecutive decisions for the Q-MRTC. Without knowing the state-transition behavior, the evaluation function is calculated by a real-time RL technique known as Q-learning [15]. After a decision is made, the consequent cost is used as an error signal feedback to the Q-MRTC to adjust the state-action pairs. Thus, the learning procedure is performed in a closed-loop iteration manner that will help the value of evaluation function converge to optimal radio resource control point.

Noticeably, the Q-function approximation is the key design issue in the implementation of the Q-learning algorithm [16], [17]. We propose to utilize a *feature extraction* method and a *radial basis function network* (RBFN) in the Q-MRTC. With the feature extraction method, the state space of the Q-function is mapped into a more compact set, which represents *resultant interference profile*. The resultant interference profile aggregates the states and, consequently, improves the convergence property. With the RBFN neural network, the storage requirement of the Q-function can be significantly reduced. Simulation results show that while keeping the QoS constraints of the packet error probability and packet transmission delay guaranteed, the Q-MRTC scheme can have higher system throughput by 80% and better users' satisfaction than the interference-based scheme [7]. Also, compared to the table lookup method, the storage requirement is reduced by 41%.

The rest of the paper is organized as follows. The system architecture and RRM are described in Section II and the design of Q-MRTC is proposed in Section III. The simulation results are presented in Section IV and the performance comparison between the Q-MRTC and interference-based schemes is also made. Finally, concluding remarks are given in Section V.

II. System Model

The physical layer and the MAC specifications for WCDMA are defined by 3GPP [18], [19]. The WCDMA has two types of uplink-dedicated physical channels (DPCHs): the uplink-dedicated physical data channel (DPDCH) and the uplink-dedicated physical control channel (DPCCH). A DPDCH is used to carry data generated by layer 2 and above and a DPCCH is used to carry layer 1 control information. Each connection is allocated a DPCH including one DPCCH and zero or several DPDCHs. The channel is defined in a frame-based structure in which the frame length $T_f = 10$ ms is divided into 15 slots with length $T_{\rm slot} = 2560$ chips, each slot corresponding to one power control period. Hence, the power control frequency is 1500 Hz. The spreading factor (SF) for DPDCH can vary between $4 \sim 256$ by $\mathrm{SF} = 256/2^k$, $k = 0, 1, \ldots, 6$, carrying 10×2^k bits per slot and the SF for DPCCH is fixed at 256, carrying 10 bits per slot. In addition, a common physical channel, called physical random-access channel (PRACH), is defined to carry uplink random-access burst(s).

Two types of services are considered in this paper: real-time service as type 1 and nonreal-time service as type 2. The system provides connection-oriented transmission for real-time traffic and best-effort transmission-rate allocation for nonreal-time traffic. To guarantee the timely constraint of real-time service, a UE always holds a DPCH while it transmits real-time packets, regardless of the variation of the required transmission rate. The real-time UE may generate variable rate information whose characteristics are indicated in its request profile. On the other hand, a UE should contend for the reservation of a DPCH to transmit a burst of nonreal-time packets and will release the DPCH immediately, while the burst of data is completely transmitted. The nonreal-time data are transmitted burst by burst.

When a UE has traffic to transmit, it first sends its service request embedded in a random-access burst via PRACH. For the service request profile, a real-time request provides the mean rate and rate variance to indicate its transmission-rate requirement, while a nonreal-time request provides the maximum and minimum rate requirements. As the base station receives the new request, the admissible transmission rate will be evaluated. Due to the service requirements, RRM performs two different kinds of decision. For a real-time request, it will be accepted or rejected. On the other hand, for a nonreal-time request, an appropriate transmission rate will be allocated. A nonreal-time request specifies the range of the required transmission rates for itself and would be blocked if the WCDMA system cannot provide a suitable transmission rate to satisfy its required transmission rate. In this paper, we assume that all packets have the same length. Also, a data packet is assumed to be transmitted in a DPDCH frame by a basic rate channel and, therefore, a multirate channel can transmit multiple data packets in a DPDCH frame.

The transmission power of a physical channel should be adjusted dependent on its spreading factor, coding scheme, rate-matching attributes, and BER requirement. Here, we assume that all physical channels adopt the same coding scheme and have the same rate-matching attributes and BER requirement. Therefore, the power allocation for a physical channel is simply dependent on its spreading factor and is in inverse proportion [20]. Since each UE determines its uplink transmission power in a distributed manner, the total received interference power at base station is time varying. For operational stability, the transmission power is determined under the consideration of maximal allowed interference power. In this way, for WCDMA systems, the SIR-based power control scheme that is specified by 3GPP is equivalent to the strength-based power control scheme. Consequently, the complexity of the multirate transmission control is reduced and the operation can disregard the variation of the received interference.

To maximize the spectrum utilization, the radio resource management is designed to accommodate as many of the access requests as possible and to allocate the transmission rate of each request as largely as possible, while the QoS requirements are fulfilled. An erroneous real-time packet will be dropped since there is no retransmission for real-time packets, while the erroneous nonreal-time packets will be recovered via the automatic repeat request (ARQ) scheme. The packet error probability (denoted by P_e) and the packet transmission delay (denoted by D_d) are considered as the system performance measures. Also, the maximum tolerable packet error probability, denoted by P_e^*, and maximum tolerable packet transmission delay time, denoted by D_d^*, are defined as the system QoS requirements.

III. Design of Q-MRTC

A. State, Action, and Transmission Cost Function

The radio resource management of a multimedia WCDMA system is regarded as a discrete-time MDP problem, where major events are arrivals of service requests in a cell. The service request arrivals would trigger the transition of the system state such that the radio resource control is executed. For the arrival of the kth request, the *system state* is assumed at x_k, defined as

$$x_k = (I_m, I_v, i, \mathbf{R}_i) \tag{1}$$

where I_m and I_v denote the mean and variance of the interference from existing connections, i indicates that x_k is an arrival of type i, and $\mathbf{R}_i$ is transmission rate requirement of the type i request, $i = 1, 2$. The (I_m, I_v) is the interference profile. Since the capacity of the WCDMA system is interference limited, the interference profile is employed to indicate the system load [3]. The $\mathbf{R}_1 = (r_m, r_v)$, where r_m and r_v denote the mean rate and the rate variance of a real-time request, respectively; the $\mathbf{R}_2 = (r_{\max}, r_{\min})$ where $r_{\max}$ and $r_{\min}$ denote the maximum rate and the minimum rate requirements of a nonreal-time request, respectively.

Based on the system state x_k, the multirate transmission controller will determine an *action* (denoted by A_k) for the kth request arrival. The action A_k is defined as

- Real-time request:

$$A_k = \begin{cases} 1 & \text{if accepted} \\ 0 & \text{if rejected} \end{cases} \tag{2}$$

- Nonreal-time request:

$$A_k = \begin{cases} r, r_{\min} \le r \le r_{\max} & \text{if accepted} \\ 0 & \text{if rejected.} \end{cases} \tag{3}$$

If the state-action pair (x_k, A_k) has been determined, an immediate transmission cost is defined as

$$(x_k, A_k) = \alpha \left[\frac{P_e(x_k, A_k) - P_e^*}{P_e^*} \right]^2 + (1-\alpha) \left[\frac{D_d(x_k, A_k) - D_d^*}{D_d^*} \right]^2 \tag{4}$$

where $P_e(x_k, A_k)$ is the packet error probability, $D_d(x_k, A_k)$ is the packet transmission delay, and α is the weighting factor. $c(x_k, A_k)$ is a random variable because channel fading and imperfect power control are not included in the state-action pair yet. We further define an evaluation function, denoted by $Q(x, A)$, as the expected total discounted cost counting from the initial state-action pair (x, A) over an infinite time. It is given by

$$Q(x, A) = E\left\{ \sum_{k=0}^{\infty} \gamma^k c(x_k, A_k) | x_0 = x, A_0 = A \right\} \tag{5}$$

where $0 \le \gamma < 1$ is a discounted factor. The multirate transmission control is to determine an optimal action, denoted by A^*, which minimizes the Q-function with respective to the current state. The minimization of the Q-function represents the maximization of the system capacity and the fulfillment of QoS requirements.

Let $P_{xy}(A)$ be the transition probability from state x with action A to the next state y. Then, $Q(x, A)$ can be expressed as

$$\begin{aligned} Q(x, A) &= E\{c(x_0, A_0) | x_0 = x, A_0 = A\} + \\ &\quad E\left\{ \sum_{k=1}^{\infty} \gamma^k c(x_k, A_k) | x_0 = x, A_0 = A \right\} \\ &= E\{c(x, A)\} + \gamma \sum_{y} P_{xy}(A) \times \\ &\quad E\left\{ \sum_{k=1}^{\infty} \gamma^{k-1} c(x_k, A_k) | x_1 = y, A_1 = B \right\} \\ &= C(x, A) + \gamma \sum_{y} P_{xy}(A) Q(y, B) \end{aligned} \tag{6}$$

where $C(x, A) = E\{c(x, A)\}$. Eq. (6) indicates that the Q function of the current state-action pair can be represented in terms of the expected immediate cost of the current state-action pair and the Q function of the next state-action pairs.

Based on the principle of Bellman's optimality [21], the optimal action A^* can be obtained by a two-step optimality operation. The first step is to find an intermediate minimal of $Q(x, A)$, denoted by $Q^*(x, A)$, where the intermediate evaluation function for every possible next state-action pair (y, B) is minimized

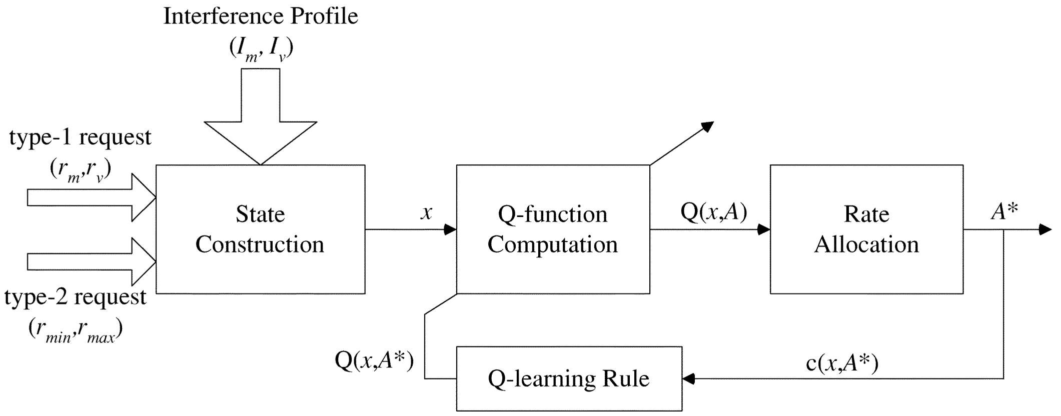


Fig. 1. Structure of Q-learning-based multirate transmission control (Q-MRTC) scheme.

and the optimal action is performed with respect to each next state y. $Q^*(x,A)$ is given by

$$Q^*(x,A) = C(x,A) + \gamma \sum_y P_{xy}(A) \left\{ \underset{B}{\text{Min}} \left[Q^*(y,B) \right] \right\} \quad \text{for all } (x,A). \tag{7}$$

Then we can determine the optimal action A^* with respect to the current state x such that $Q^*(x,A)$ is minimal, which can be expressed as

$$Q^*(x,A^*) = \underset{A}{\text{Min}} \left[Q^*(x,A) \right]. \tag{8}$$

However, it is difficult to find the $C(x,A)$ and $P_{xy}(A)$ to solve (7). In this paper, we adopt a real-time reinforcement learning algorithm, called the Q-learning algorithm [15], [16], to find the optimal resource allocation without *a priori* knowledge of $C(x,A)$ and $P_{xy}(A)$. To find the optimal $Q^*(x,A)$, the Q-learning algorithm computes the Q value in a recursive method using available information $(x,\ A,\ y,\ c(x,A))$, where x and y are the current and the next states, respectively, and A and $c(x,A)$ are the action for current state and its immediate cost of the state action pair, respectively.

B. Q-MRTC

Fig. 1 shows the structure of the Q-learning-based multirate transmission control (Q-MRTC) scheme. When a service request arrives at system state x, the Q-function computation block computes the value of $Q(x,A)$ for every possible action A. The *rate allocation* block then determines the optimal rate allocation A^* or call rejection with respect to all the current Q values of all possible actions. In the *Q-learning-rule* block, the immediate cost $c(x,A^*)$ can be observed and the Q-learning rule is used to adjust the value of $Q(x,A)$. The Q-learning rule is formulated by

$$Q(x,A) = \begin{cases} Q(x,A) + \eta \Delta Q(x,A), & \text{if } A = A^* \\ Q(x,A) & \text{otherwise} \end{cases} \tag{9}$$

where η is the learning rate $0 \le \eta \le 1$ and

$$\Delta Q(x,A^*) = \left\{ c(x,A^*) + \gamma \underset{B}{\text{Min}} \left[Q(y,B) \right] \right\} - Q(x,A^*). \tag{10}$$

Since only one action pair is chosen for evaluation in each learning epoch, for the Q-learning rule, only the Q value of the chosen action pair is updated, while others are kept unchanged. Also, in (10), the operation of $\underset{B}{\text{Min}} \left[Q(y,B) \right]$ is executed by comparing the Q values of all the possible action candidates for state y and then choosing the desired action B with minimal Q value.

In [15], Watkins and Dayan had proved the convergence theorem of Q-learning. Here, the theorem is restated as: *if the value of each admissible pair is visited infinitely often and the learning rate is decreased to zero in a suitable way, then the value of* $Q(x,A)$ *in (9) will converge to* $Q^*(x,A)$ *with probability 1.*

Usually, if the state space is too large, it would require a huge amount of memory to store the values of Q-function and take a long time for the Q-learning algorithm to converge. To tackle the above problems, in the Q-function computation block, we employ the *feature extraction* method and *radial basis function network* (RBFN) for the Q-function approximation in the proposed Q-MRTC. Here, the state-action pair is first transformed into a dimension-reduced feature vector; then, the feature vector is used as input parameters to compute the corresponding Q value that is stored in the RBFN network.

The feature extraction method maps the original state-action pair into a feature vector that must be properly chosen to reflect the important behavior characteristics of the state-action pair [17]. In the WCDMA system, after the state action is performed, the change of interference is the most significant outcoming response. Therefore, the feature vector of (x,A) is selected to be the *resultant interference profile*, denoted by $(I_m + \Delta I_m, I_v + \Delta I_v)$, where $(\Delta I_m, \Delta I_v)$ indicates the change of interference profile (I_m, I_v) due to action A at state x. In other words, the state-action pair (x,A) can be converted to resultant interference profile $(I_m + \Delta I_m, I_v + \Delta I_v)$. It is noted that the dimension of the resultant interference profile is smaller than that of the original state-action pairs. While a strength-based closed-loop power control is assumed, the received power for a unit of transmission rate is set to 1. Consequently, $(\Delta I_m, \Delta I_v)$ is obtained by

$$(\Delta I_m, \Delta I_v) = \begin{cases} (r_m, r_v) & \text{if accepts a real-time request,} \\ (r, 0) & \text{if accepts a nonreal-time request with rate } r, \\ (0,0) & \text{if rejects a request.} \end{cases} \tag{11}$$

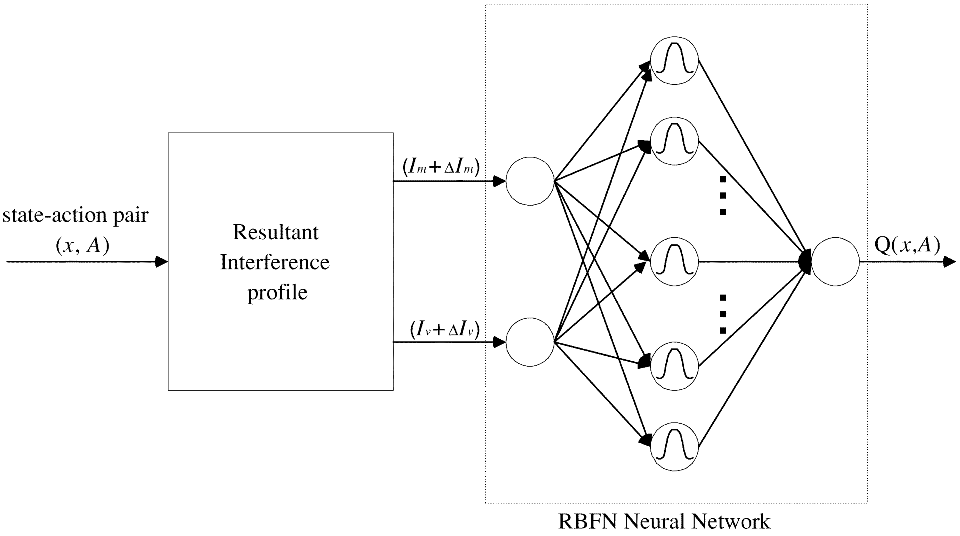


Fig. 2. Q-function computation by RBFN neural network.

RBFN is a three-layer self-growing neural network, including an input layer, an output layer, and a hidden layer [16]. The hidden layer consists of a sequence of nodes whose activation functions are normalized Gaussian. The RBFN neural network performs a function approximation for the Q function. When the RBFN is well trained, the Q values of all the state-action pairs are stored in the RBFN. With the input parameters of the resultant interference profile, the RBFN calculate the corresponding Q value.

The key concept of RBFN is *local tuning* and *separated storage*. Each node in the hidden layer represents a part of the characteristics of the input vectors and stores these characteristics locally. Thus, it breaks a large dimensional mapping function into multiple small-dimensional functions. Due to the separated storage property, only some hidden nodes in the RBFN would be adjusted with respect to the new input error signal, which can reduce the training epoch significantly. Fig. 2 shows the Q-function computation performed by the RBFN. The state-action pair (x, A) is mapped into its corresponding resultant interference profile $(I_m + \Delta I_m, I_v + \Delta I_v)$ and the RBFN neural network then calculates $Q(x, A)$ as a function of $(I_m + \Delta I_m, I_v + \Delta I_v)$. The well-known back-propagation learning rule is applied in the training process.

The Q value for state-action pair (x, A) is updated by (9) when the next request arrives and $\eta \Delta Q(x, A)$ is served as an error signal, which is backpropagated in the neural network. With the feature extraction method and RBFN neural network, the Q-MRTC can obtain $Q(x, A)$ efficiently through the online operation. As noted, $Q(x, A)$ will approach to $Q^*(x, A)$ through the training procedure, while the convergence theorem of Q-learning holds.

C. Parameter Initialization

Before the Q-MRTC is performed for the online operation, it is necessary to assign a proper set of initial values. An appropriate initialization can provide a good relationship of the input parameters and the decision output for an event at the beginning of system operation such that the transient period of Q-learning procedure would be short. To obtain the initial Q values, the composite interference received at base station is assumed to be log-normally distributed. Although the assumption of log-normal distribution may not hold in some cases, it indeed provides a meaningful initial guess rather than a random initialization.

For a given state-action pair (x, A), the initial value of $Q(x, A)$ is set according to QoS measurements. Since the packet transmission delay cannot be calculated in advance, the normalized expected packet error probability is preferred as the initial value of $Q(x, A)$ and is expressed as $\left(\overline{P}_e(x, A) - P_e^*)^2/P_e^*\right)$, where $\overline{P}_e(x, A)$ is the expected packet error probability if the state-action pair (x, A) is performed. The $\overline{P}_e(x, A)$ is given by

$$\overline{P}_e(x, A) = 1 - \left(1 - \int P_b(I)\mathcal{L}(I)dI\right)^L \qquad (12)$$

where L is the packet length, $P_b(I)$ is the bit error probability at the interference level I, and $\mathcal{L}(I)$ is the log-normal function for interference level I with mean $(I_m + \Delta I_m)$ and variance $(I_v + \Delta I_v)$. The $P_b(I)$ is given by [5]

$$P_b(I) = \frac{\kappa \exp -\beta * G}{I} \qquad (13)$$

with parameters of κ and β, which are adjustable for matching with a particular coding scheme, and G is the spreading factor of a basic rate channel.

In summary, the procedure of Q-MRTC is described as follows.

- **Step 1:** *State-Action Construction*
  Construct the current state $x = (I_m, I_v, i, \mathbf{R}_i)$ and find a set of all possible actions for state x, denoted by $\mathbf{A}(x)$, when a new request arrives.
- **Step 2:** *Q-Value Computation*
  For the set of state-action pairs $\{(x, A) \mid A \in \mathbf{A}(x)\}$, compute the respective $Q(x, A)$ values by the RBFN neural network.

- **Step 3:** *Rate Allocation*

  Determine the optimal action A^* such that the value of $Q(x, A^*)$ is minimum, i.e., $Q(x, A^*) = \underset{A \in \mathbf{A}(x)}{\mathrm{Min}} [Q(x, A)]$.
- **Step 4:** *Q-Value Update*

  Update the Q values by (9) as the next event arrives with state y and the online cost $c(x, A^*)$ is obtained. Since the Q value is stored in a neural network, $\eta \Delta Q(x, A^*)$ is used as an error signal backpropagated into the neural network, instead of the error between the desired and the actual outputs. Go to Step 1.

IV. Simulation Results and Discussion

In this simulation, two kinds of traffic are transmitted via the real-time service: one is two-level transmission rate traffic and the other is M-level transmission rate traffic. They are modeled by two- and M-level Markov modulated deterministic process (MMDP), respectively. The two-level MMDP is generally used to formulate ON–OFF voice traffic stream and the M-level MMDP is to formulate the advanced speech or other real-time traffic streams, e.g., video. On the other hand, the nonreal-time service is considered to transmit variable-length data bursts. The arrival process of the data burst is Poisson and the data length is assumed to be with a geometric distribution. A data burst can carry any type of wireless data, e.g., e-mail, wireless markup language (WML) pages, etc. The detailed traffic parameters are listed in Table I. A basic rate in the WCDMA system is assumed to be a physical channel with $SF = 256$. For each connection, DPCCH is always active to maintain the connection reliability. To reduce the overhead cost of interference produced by DPCCHs, the transmitting power of a DPCCH is smaller than its respective DPDCH by an amount of 3 dB. The other simulation parameters are given as $P_e^* = 0.01$, $D_d^* = 0.5\ s$, and $\alpha = 0.3$.

A conventional interference-based scheme proposed in [7] is used as a benchmark for comparison with Q-MRTC. The interference-based scheme would admit the connection for a real-time request or allocate a transmission rate for a nonreal-time request if the expected packet error probability in terms of the resultant SIR is smaller than the QoS requirement.

Fig. 3 illustrates the throughput of the Q-MRTC and the interference-based scheme versus the request arrival rate. The Q-MRTC has throughput higher than the interference-based scheme and the throughput improvement becomes greater as the request arrival rate becomes larger. Generally speaking, Q-MRTC can improve the maximum throughput by an amount of 80% over the interference-based scheme. The reason is that, in the Q-MRTC, the transmission cost comprises the cost of immediate and consecutive decision and the behavior of interference variation is taken into consideration for multirate transmission control. Also, the Q-MRTC performs an online reinforcement learning algorithm to estimate the transmission cost. The estimation error is backpropagated to the Q-MRTC and reduced through the closed-loop learning procedure. Therefore, the Q-MRTC could provide a more accurate estimation for multirate transmission cost and greater throughput improvement when the traffic load becomes large. On the other hand, the interference-based scheme generally estimates the multirate transmission cost of packet error probability at the instant of a request arrival. Actually, some existing connections may terminate or handoff between two consecutive arrivals and the received interference level decreases subsequently. Therefore, the interference-based scheme would overestimate the multirate transmission cost.

TABLE I
Traffic Parameters in the Multimedia WCDMA System

Traffic Type	Traffic Parameters
2-level real-time	Call holding time: 30 seconds Mean talkspurt duration: 1.00 seconds Mean silence duration: 1.35 seconds
M-level real-time	Call holding time: 30 seconds Peak rate (M): 4-fold of basic rate Mean rate: 2-fold of basic rate
Non-real-time	Mean data burst size: 200 packets $r_{\min}$: 1-fold of basic rate $r_{\max}$: 8-fold of basic rate

Fig. 4 illustrates the blocking probability versus the request arrival rate. It can be found that the blocking probability of the Q-MRTC is much smaller than that of the interference-based scheme for real-time and nonreal-time requests and that the blocking probabilities of the real-time requests are higher than those of the nonreal-time requests. The reason is that the admitted transmission rate of the nonreal-time requests are negotiable. It can also be seen that Q-MRTC has a larger difference between the real-time and nonreal-time blocking probabilities than the interference-based scheme. This is because the interference-based scheme generally accommodates fewer connections and operates in a lower interference condition so that the interference variation due to the variable-rate transmission behavior of the real-time requests is smaller. By contrast, Q-MRTC accommodates more requests and operates in a higher interference situation so that the interference variation produced by the real-time requests becomes more critical. That is, the variable-rate transmission behavior contributes a higher admission cost for the Q-MRTC.

We further define an overall users' satisfaction index (USI)m which is a linear combination of A_{a1}/A_{d1} (type 1) and A_{a2}/A_{d2} (type 2), where the A_{a1} (A_{a2}) is the admitted transmission rate for type 1 (type 2) and the A_{d1} (A_{d2}) is the desired transmission rate for type 1 (type 2); $A_{d1} = 1$ and $A_{d2} = r_{\max}$. That is, USI is expressed as

$$\mathrm{USI} = \gamma \frac{A_{a1}}{A_{d1}} + (1 - \gamma) \frac{A_{a2}}{A_{d2}} \tag{14}$$

where γ is the weighting factor.

Fig. 5 depicts USI versus the request arrival rate for different traffic patterns, where P_{RT}, denoting the percentage of the real-time traffic-arrival requests in the traffic load, varies from 0.1 to 0.3. It can be found that Q-MRTC has higher USI than the interference-based scheme and the improvement is more significant as the traffic load becomes heavier. This is because Q-MRTC can accurately estimate the multirate transmission cost. Also, USI decreases as the request arrival rate increases. Since the high traffic load may decrease the admitted transmission rate for new requests, the USI value

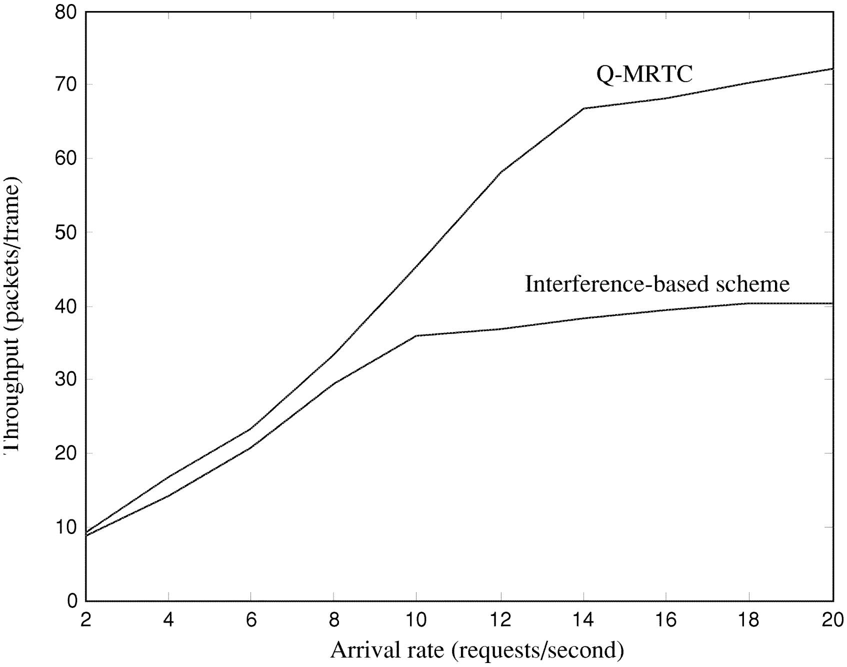


Fig. 3. Throughput versus the request arrival rate.

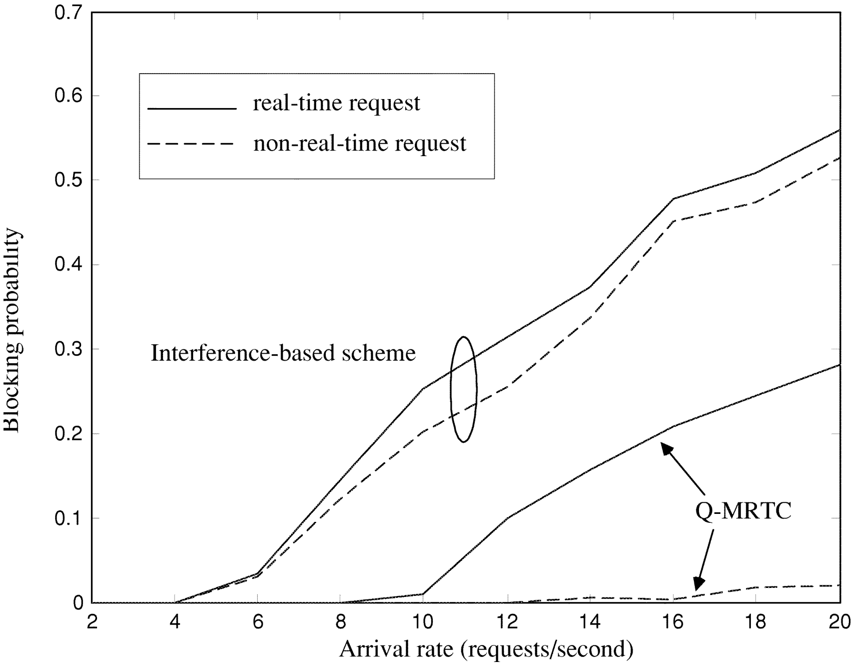


Fig. 4. Blocking probability versus request arrival rate.

decreases consequently. Another observation is that, under the fixed weighting factor ($\alpha = 0.3$), the USI decreases as P_{RT} increases. This is because the real-time requests produce interference variation higher than nonreal-time ones do, which leads to larger real-time blocking probability and less nonreal-time admitted transmission rate.

Fig. 6 depicts USI versus the request arrival rate for different weighting factors $\gamma = 0.3$, 0.5, and 0.7. It can be found that the USI of Q-MRTC is lower when γ is larger (more weighting on type-1 service) because Q-MRTC accommodates more requests and operates under higher interference condition. Thus, the interference variation produced by real-time requests becomes critical. From Figs. 5 and 6, it can be concluded that the variable-rate transmission characteristic of real-time requests plays an important role for the multirate transmission control in the multimedia WCDMA systems.

Fig. 7 shows the QoS measures 1) the packet error probability and 2) the packet transmission delay versus the request

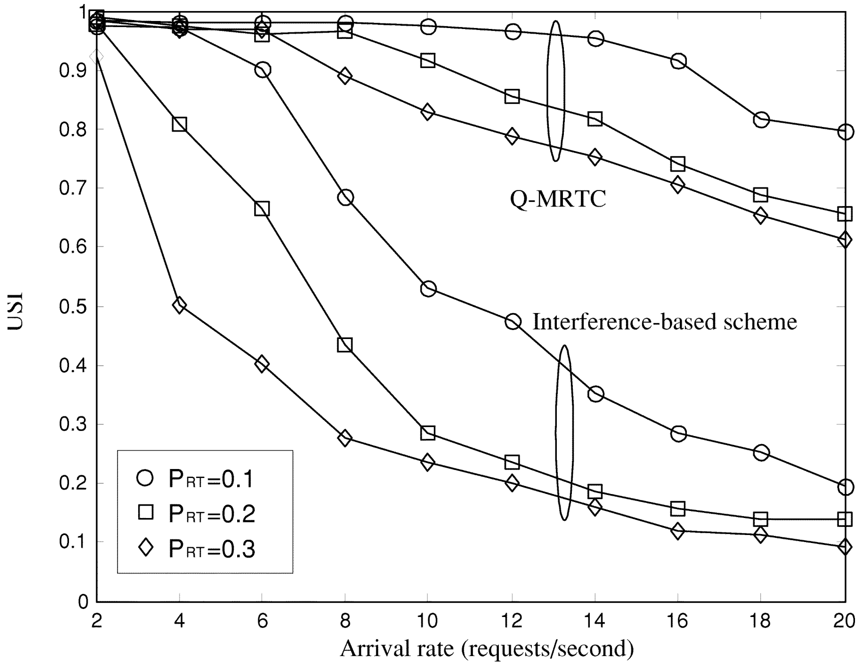


Fig. 5. USI versus the request arrival rate for different traffic patterns.

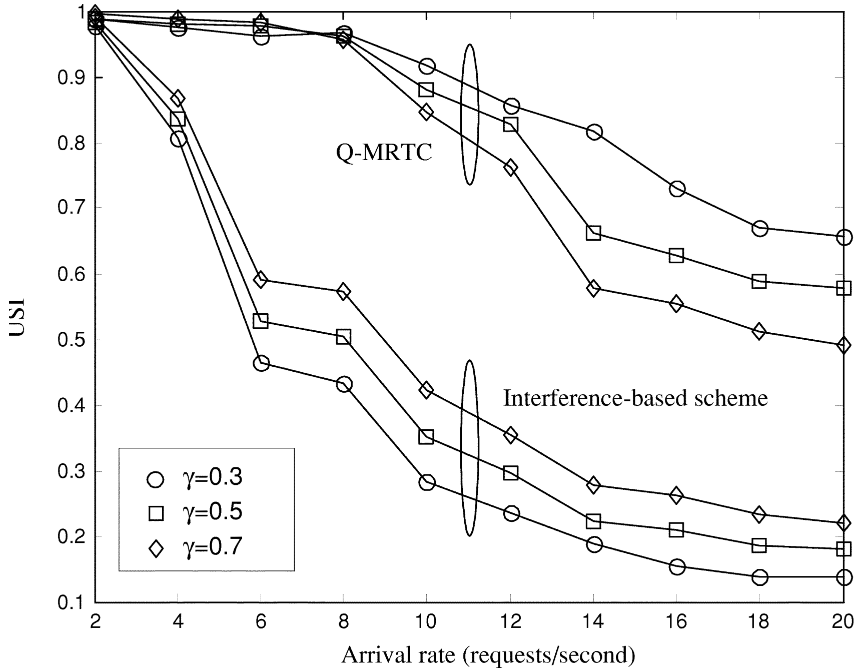
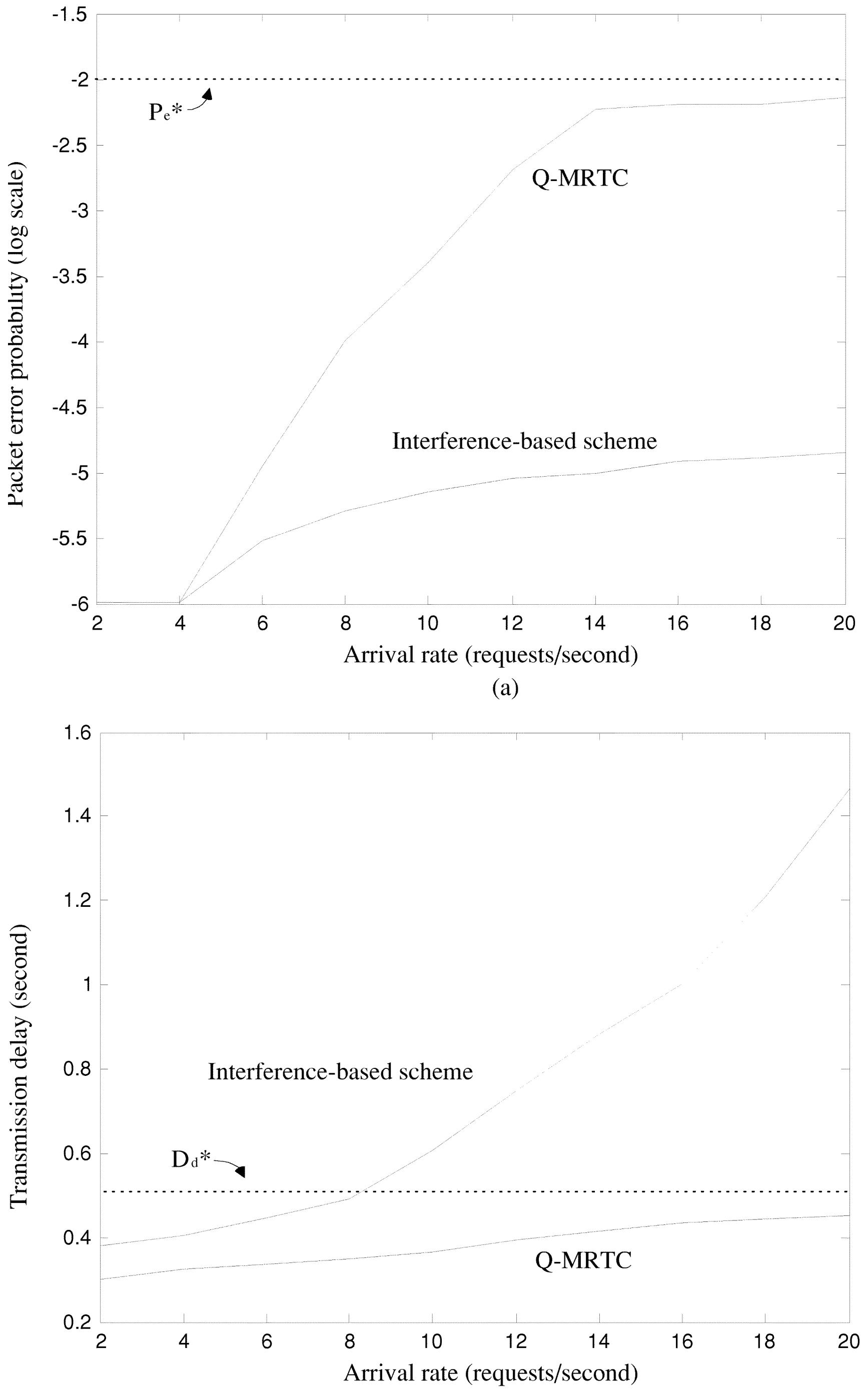


Fig. 6. USI versus the request arrival rate for different weighting factors.

arrival rate. It can be seen that Q-MRTC can always keep the QoS requirements of packet error probability and packet transmission delay. By contrast, only the QoS requirement of packet error probability is kept in the interference-based scheme. It is because Q-MRTC dynamically evaluates the transmission cost that is in terms of packet error probability and packet transmission delay. The Q-MRTC is more suitable for multimedia WCDMA systems than is the interference-based scheme. Also, it can be seen that the average packet error probability of the Q-MRTC is larger than that of the interference-based scheme; however, the Q-MRTC can still hold the packet error probability within the QoS constraint. This is because the interference-based scheme is too conservative in the multirate transmission control and because it admits less requests and allocates lower transmission rates. On the other hand, the Q-MRTC obtains the transmission cost from the online operation of the WCDMA system. Consequently, it can accommodate more requests and appro-



(a)

(b)

Fig. 7. QoS measures. (a) Packet error probability and (b) transmission delay versus the request-arrival rate.

priately allocate transmission rates as much as possible, under the QoS constraints.

To evaluate the performance of storage requirement reduction, we assume a table lookup method in which the continuous-valued parameters of the resultant interference profile are partitioned into several discrete levels. Generally, a different number of discrete levels leads to different system throughput and different storage requirements. As an example for comparison, we divide the interference mean $(I_m + \Delta I_m)$ of the resultant interference profile into 40 levels and the interference variance $(I_v + \Delta I_v)$ into 10 levels, which has similar system performance as a RBFN neural network. Table II shows the number of required storage units. There are 400 storage units required in the table lookup method. On the contrary, only 118 hidden neuron nodes are required in the RBFN neural network. While there are two parameters in each hidden node, there are 236 storage units required for RBFN. Therefore, RBFN can achieve storage requirement reduction of 41%. Furthermore, the table lookup method is static partioned and it is hard to find a proper partition level, especially for bursty traffic. However, in the RBFN neural network, the value of each meaningful state-action pair is stored and adjusted separately in a corre-

TABLE II
NUMBER OF REQUIRED STORAGE UNITS: AN EXAMPLE

Method	Storage Units
Table lookup	400
RBFN	236 (118 hidden nodes)

sponding hidden node. That is, the storage space is nonlinearly partitioned in the RBFN neural network. While the traffic load and pattern change with time, the hidden nodes of the RBFN neural network can self-organize dynamically and the storage space can be repartitioned accordingly.

V. Concluding Remarks

In this paper, we propose a Q-learning-based multirate transmission control scheme for radio resource management in multimedia WCDMA systems. The Q-learning algorithm is applied to accurately estimate the transmission cost for the multirate transmission control and the feature extraction method and radial basis function network are employed for Q-function approximation that maps the original state-action pairs into the *resultant interference profile*.

Simulation results show that Q-MRTC can improve the throughput of multimedia WCDMA system by 80% over the conventional interference-based scheme proposed in [7], under the constraint of the QoS requirements of packet error probability and packet transmission delay. Also, the Q-MRTC provides better users' satisfaction. It is because the Q-learning algorithm performs closed-loop control by applying the system performance measures as a feedback to adjust the multirate transmission cost and, correspondingly, the Q-MRTC can have self-tuning capability to adaptively estimate the transmission cost. Moreover, the storage requirement of RBFN neural network is less than that of the conventional table-lookup method by the amount of 41%.

The multirate transmission control considered in this paper is only at the call/burst level. However, since the interference profile may change during the call/burst holding time, it is possible for the system, during the service holding time, to renegotiate/reallocate the transmission rate according to the variation of interference profile. Combining a real-time scheduling algorithm with the Q-MRTC to further enhance the communication quality and achieve higher system throughput can be studied further.

Acknowledgment

The authors would like to thank the anonymous reviewers for their comments, which improved the correctness and presentation of the paper.

References

[1] S. Choi and K. G. Shin, "An uplink CDMA system architecture with diverse QoS guarantees for heterogeneous traffic," *IEEE/ACM Trans. Networking*, vol. 7, pp. 616–628, Oct. 1999.

[2] K. S. Gilhousen, I. M. Jacobs, R. Padovani, A. J. Viterbi, L. A. Weaver, and C. E. Wheatley, "On the capacity of a cellular CDMA system," *IEEE Trans. Veh. Technol.*, vol. 40, pp. 303–312, May 1991.

[3] W. Huang and V. K. Bhargava, "Performance evaluation of a DS-CDMA cellular system with voice and data services," in *Proc. IEEE PIMRC'96*, 1996, pp. 588–592.

[4] A. Hämäläinen and K. Valkealahti, "Adaptive power increase estimation in WCDMA," in *IEEE PIMRC'02*, vol. 3, Sept. 2002, pp. 1407–1411.

[5] J. B. Kim and M. L. Honig, "Resource allocation for multiple classes of DS-CDMA traffic," *IEEE Trans. Veh. Technol.*, vol. 49, pp. 506–519, Mar. 2000.

[6] S. M. Shin, C.-H. Cho, and D. K. Sung, "Interference-based channel assignment for DS-CDMA cellular systems," *IEEE Trans. Veh. Technol.*, vol. 48, pp. 233–239, Jan. 1999.

[7] N. Dimitriou and R. Tafazolli, "Quality of service for multimedia CDMA," *IEEE Commun. Mag.*, vol. 38, pp. 88–94, July 2000.

[8] V. Phan-Van and D. Luong, "Capacity enhancement with simple and robust soft-decision call admission control for WCDMA mobile cellular PCNs," in *Proc. IEEE Vehicular Technology Conf. (VTC'01)*, vol. 3, Oct. 2001, pp. 1349–1353.

[9] D. Mitra, M. I. Reiman, and J. Wang, "Robust dynamic admission control for unified cell and call QoS in statistical multiplexers," *IEEE J. Select. Areas. Commun.*, vol. 16, pp. 692–707, June 1998.

[10] K. W. Ross, *Multiservice Loss Models for Broadband Communication Networks*. Berlin, Germany: Springer-Verlag, 1995.

[11] J. Nie and S. Haykin, "A Q-learning-based dynamci channel assignment technique for mobile communication systems," *IEEE Trans. Veh. Technol.*, vol. 48, pp. 1676–1687, Sept. 1999.

[12] H. Tong and T. X. Brown, "Adaptive admission call admission control under quality of service constraints: A reinforcement learning solution," *IEEE J. Select. Areas. Commun.*, vol. 18, pp. 209–221, Feb. 2000.

[13] B. Makarevitch, "Application of reinforcement learning to admission control in CDMA network ," in *Proc. IEEE PIMRC '00*, pp. 1353–1357.

[14] P. Marbach, O. Mihatsch, and J. N. Tsisiklis, "Call admission control and routing in integrated services networks using neuro-dynamic programming," *IEEE J. Select. Areas. Commun.*, vol. 18, pp. 197–208, Feb. 2000.

[15] C. J. C. H. Watkins and P. Dayan, "Q-learning," *Machine Learning*, vol. 8, pp. 279–292, 1992.

[16] S. Haykin, *Neural Networks*, 2nd ed. Englewood Cliffs, NJ: Prentice-Hall, 1999.

[17] D. P. Bertsekas and J. N. Tsitsiklis, *Neuro-Dynamic Programming*. Belmont, MA: Athena Scientific, 1996.

[18] 3rd Generation Partnership Project. (1999, Dec.) Physical Channels and Mapping of Transport Channels onto Physical Channels (FDD). *3GPP TS25.211* [Online]http://www.3gpp.org

[19] 3rd Generation Partnership Project. (1999, Dec.) MAC Protocol Specification. *3GPP TS25.321* [Online]http://www.3gpp.org

[20] 3rd Generation Partnership Project. (2001, Mar.) Physical Layer Procedures (FDD). *3GPP TS25.214* [Online] http://www.3gpp.org

[21] R. Bellman, *Dynamic Programming*. Princeton, NJ: Princeton Univ. Press, 1957.

Yih-Shen Chen (S'00) was born in Miaoli, Taiwan, in September 1973. He received the B.E. and M.E. degrees in communication engineering from National Chiao-Tung University, Hsinchu, Taiwan, in 1995 and 1997, respectively. He is currently working toward the Ph.D. degree at the same university.

From 1997 to 1999, he was an Engineer in the Protocol Design Technology Department, Computer and Communications Research Laboratories, Industrial Technology Research Institute, Hsinchu, Taiwan, where he was involved in the design of software protocol for the DECT networks. His research interests include performance analysis, protocol design, and mobile radio networks.

Chung-Ju Chang (S'84–M'85–SM'94) was born in Taiwan, R.O.C., in August 1950. He received the B.E. and M.E. degrees in electronics engineering from National Chiao Tung University (NCTU), Hsinchu, Taiwan, in 1972 and 1976, respectively, and the Ph.D degree in electrical engineering from National Taiwan University (NTU), Taipei, Taiwan, in 1985.

From 1976 to 1988, he was with Telecommunication Laboratories, Directorate General of Telecommunications, Ministry of Communications, Taiwan, as a Design Engineer, Supervisor, Project Manager, and then Division Director. There, he was involved in designing digital switching system, RAX trunk tester, ISDN user-network interface, and ISDN service and technology trials in Science-Based Industrial Park. In the meantime, he also acted as a Science and Technical Advisor for the Minister of the Ministry of Communications from 1987 to 1989. In 1988, he joined the Faculty of the Department of Communication Engineering and the Center for Telecommunications Research, College of Electrical Engineering and Computer Science, National Chiao Tung University, as an Associate Professor. He has been a Professor since 1993. He was Director of the Institute of Communication Engineering from August 1993 to July 1995 and Chairman of the Department of Communication Engineering from August 1999 to July 2001. He is now the Dean of the Research and Development Office, NCTU. His research interests include performance evaluation, wireless communication networks, and broad-band networks.

Dr. Chang is a Member of the Chinese Institute of Engineers (CIE) and has been an Advisor for the Ministry of Education to promote the education of communication science and technologies for colleges and universities in Taiwan since 1995. He is also a Committee Member of the Telecommunication Deliberate Body.

Fang-Chin Ren (S'94–A'01–M'01) was born in Hsinchu, Taiwan. He received B.E., M.E., and Ph.D. degrees in communication engineering from National Chiao Tung University (NCTU), Hsinchu, Taiwan, in 1992, 1994, and 2001, respectively.

Since 2001, he has been an Engineer in the Protocol Design Technology Department, Computer and Communications Research Laboratories, Industrial Technology Research Institute, Hsinchu. He is currently involved in the design of software for protocol and hardware integration on the third-generation wireless communications. His current research interests include performance analysis, protocol design, and mobile radio network.