Copyright © 2008 IEEE. Reprinted from IEEE Transactions on Automatic Control, 2007; 52 (8):1442-1448

This material is posted here with permission of the IEEE. Such permission of the IEEE does not in any way imply IEEE endorsement of any of the University of Adelaide's products or services. Internal or personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution must be obtained from the IEEE by writing to pubs-permissions@ieee.org.

By choosing to view this document, you agree to all provisions of the copyright laws protecting it.

Technical Notes and Correspondence

Interference-Based Dynamic Pricing for WCDMA Networks Using Neurodynamic Programming

Siew-Lee Hew and Langford B. White

***Abstract*—We study the problem of optimal integrated dynamic pricing and radio resource management, in terms of resource allocation and call admission control, in a WCDMA network. In such interference-limited network, one's resource usage also degrades the utility of others. A new parameter *noise rise factor*, which indicates the amount of interference generated by a call, is suggested as a basis for setting price to make users accountable for the congestion externality of their usage. The methods of dynamic programming (DP) are unsuitable for problems with large state spaces due to the associated "curse of dimensionality." To overcome this, we solve the problem using a simulation-based neurodynamic programming (NDP) method with an action-dependent approximation architecture. Our results show that the proposed optimal policy provides significant average reward and congestion improvement over conventional policies that charge users based on their load factor.**

***Index Terms*—Dynamic pricing, neurodynamic programming (NDP), radio resource management (RRM), WCDMA.**

I. Introduction

CDMA-based systems like WCDMA have soft capacity, where their effective capacity is not determined by the available resources as in TDMA. Each user experiences interference from users outside its cell, in addition to the ones within the same cell. Good interference handling via radio resource management (RRM) plays an important role in increasing system capacity and providing Quality of Service (QoS) guarantee. Dynamic pricing has been proposed as a mechanism to encourage users to adapt their resource consumption level according to network conditions. A good dynamic pricing model can provide the necessary positive incentives to increase users' arrival rate when the network load is relatively low and negative incentives for users to defer their usage when the load is relatively high. Dynamic pricing also enhances operators' ability to recover costs and make profits to finance capacity expansions. By influencing the demand patterns, operators could avoid the costly need to provision a network so that it can always meet its peak demand.

Many dynamic pricing models have been proposed for fixed-capacity networks. The major proposals are auction-based, smart market pricing [1], shadow pricing [2], and stochastic congestion pricing [3], [4]. We refer our readers to [5] for a comprehensive survey on the pricing schemes for fixed-capacity network. In the case of wireless networks, the bulk of the pricing literature is motivated by power control using noncooperative game theory. The common approach is to first define a suitable user utility function, for example, in terms of throughput per terminal life [6], [7] the sigmoid function [8], [9] and the step function [10], [11] of signal-to-interference ratio (SIR). Users then enter into a decentralized, noncooperative game to select the transmission power that maximizes their utility based on a price announced by the base station. However, pricing is only used as an *internal* control mechanism and does not reflect the actual prices users would end up paying. Therefore, it is not clear how users would respond to the "price" proposed. These schemes also fail to utilize price as a positive incentive to encourage usage when the network is lightly loaded since the price used is only a static parameter that is designed to control the usage of, and optimized for, a fixed number of *existing* users in the network. The outcome of noncooperative games, i.e., the Nash equilibrium, is well known to be inefficient even with the introduction of pricing.

In this correspondence, we study the problem of optimal integrated dynamic pricing and RRM, in terms of call admission control (CAC) and resource allocation, in a multiservice WCDMA network. A major difference with much of the pricing literature for wireless networks is our proposal to charge users based on their contribution of the interference in a network, measured by their noise rise factor. We verify that a user's transmission rate has nonlinear effects on the network's noise rise as the system load factor approaches its maximum threshold in CDMA-based networks. In such networks, where one's usage degrades the utility of others, users should be made accountable for their congestion externality by charging them the direct and external costs of their usage. We call pricing strategies of this nature interference-based pricing. By contrast, current load-based pricing schemes for wireless networks fail to capture such externality. We will verify, using simulations, that such interference-based schemes provide significant reward and congestion improvements, compared to load-based schemes.

Unlike [6]–[11], which have been developed to deal with static scenarios and optimized for a fixed number of users, our model builds on the works of [3], [4] to allow network operators to shape demand based on their knowledge of the stochastic nature of users' arrivals and departures. By considering dynamic user arrivals, handoffs and departures, the objective is to develop an integrated optimal policy that maximizes the long-term, expected reward. Handoff call dropping is minimized via CAC. Our approach also exploits another characteristic of WCDMA services, *viz.* that it can operate within a range of transmission rates. For example, the UMTS adaptive multirates (AMR) voice codec offers transmission rates that vary between 4.75 and 12.2 kbit/s for conversational voice service [12]. As the level of interference increases, the network recalculates the optimal transmission rate of all services such that existing connections can be maintained with the addition of new users. Therefore, a service can be further classified according to the range of acceptable transmission rates, which reflects users' perception of QoS. Unlike typical congestion-dependent pricing schemes that increase price as the number of users increase, our model allows the possibility of maintaining the price if existing users are tolerant towards the degradation of QoS during their call and the long-term, expected reward of the operator is still maximized.

This problem is naturally formulated as a dynamic programming (DP) problem, but the evaluation function is too complex for an exact solution. Offline DP methods are of limited utility for problems with large state spaces because they require full expansion of all possible states and storing the reward for each state. This often leads to space

Manuscript received October 18, 2005; revised September 5, 2006. Recommended by Associate Editor I.-J. Wang. This work was supported by the CRC for Smart Internet, Australia.

S.-L. Hew is with the FB Rice & Co, Sydney, 2000 N.S.W., Australia (e-mail: shew@fbrice.com.au).

L. White is with the School of Electrical and Electronic Engineering, The University of Adelaide, Adelaide, 5005 S.A., Australia (e-mail: lwhite@eleceng.adelaide.edu.au).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TAC.2007.902746

0018-9286/$25.00 © 2007 IEEE

Authorized licensed use limited to: IEEE Xplore. Downloaded on October 23, 2008 at 20:27 from IEEE Xplore. Restrictions apply.

complexity exponential in the number of state variables, the situation infamously known as the "curse of dimensionality." We will use neurodynamic programming (NDP) [13], a simulation-based learning method, to solve the problem. This method has been successfully applied in a CAC problem [14] and a retailer inventory management problem in [15]. The rest of the paper is organized as follows. In Section II, we describe our network model. DP and NDP problems are formulated in Section III and Section IV, respectively. Experimental results are presented in Section V and conclusions in Section VI.

II. System Model

We consider the uplink of a WCDMA system with J services. New and handoff calls of class j arrive at the cell according to Poisson process with rates λ_j^{n} and λ_j^{h}, respectively. The call holding time and cell residence time of a class j call are both exponentially distributed with mean $1/\mu_j$ and $1/\gamma_j$, respectively. During a connection, it is assumed that a call alternates between ON and OFF states at rate α_j and β_j. We denote the probability that a connection is active as the activity factor $\nu_j = \beta_j/(\alpha_j + \beta_j)$. Although the system cannot distinguish between ON and OFF periods, idle periods do not contribute any interference.

Users arrive with a mean budget or willingness to pay (WTP) of Ψ_j that quantifies the satisfaction gained from a call. The system state can be represented by a vector, $\mathbf{n} = (n_1, \ldots, n_J)$, where n_j is the number of admitted users of service j. The state space of the system is finite, albeit very large, and depends on the interference generated by users within and outside of the cell. The controller jointly controls the price and RRM strategy of the system by computing and exercising the optimal integrated policy $\mathbf{u} = (u_{\mathrm{p}}(\mathbf{n}), \mathbf{u}_{\mathrm{c}}(\mathbf{n}), \mathbf{u}_{\mathrm{r}}(\mathbf{n}))$, where $u_{\mathrm{p}}(\mathbf{n})$, $\mathbf{u}_{\mathrm{c}}(\mathbf{n})$, $\mathbf{u}_{\mathrm{r}}(\mathbf{n})$ are the state-dependent admission price, call admission actions and transmission rate vector of *existing* users. In WCDMA, services can operate within a range of transmission rates. The transmission rate of service j users belongs to a finite set $\mathcal{R}_j = \{R_{j1}, \ldots, R_{jM_j}\}$, where M_j is the number of discrete transmission rates supported by the system. The discretisation of transmission rate is due to the allocation of codes. Therefore, the state-dependent transmission rate vector can be further defined as $\mathbf{u}_{\mathrm{r}}(\mathbf{n}) = (u_{\mathrm{r}_1}, \ldots, u_{\mathrm{r}_J})$, $u_{\mathrm{r}_j} \in \mathcal{R}_j$.

In order for a signal to be received, the ratio of its received power to the sum of the background noise and interference must be greater than a given target. When there are $\mathbf{n}$ users transmitting simultaneously in a given cell, the target quality is translated to the following inequality that must be satisfied for each user $i = 1, \ldots, n_j$ of service $j = 1, \ldots, J$ [12], [16]

$$\frac{W}{\nu_j u_{\mathrm{r}j}(\mathbf{n})} \times \frac{P_j}{P_{\mathrm{N}} + I_{\mathrm{own}} + I_{\mathrm{other}} - P_j} \geq \left(\frac{E_b}{N_0}\right)_j \qquad (1)$$

where W is the WCDMA chip rate, ν_j is the activity factor, $u_{\mathrm{r}j}(\mathbf{n})$ is the allocated transmission rate, P_j is the received signal power from the ith user, P_{N} is the background thermal noise power, I_{other} and I_{own} are the other-cell and own-cell interference, and $(E_b/N_0)_j$ is the ratio of energy per bit to noise density required to meet a predefined bit error rate. The total received interference at the Node B is defined as $I_{\mathrm{total}} = P_{\mathrm{N}} + I_{\mathrm{own}} + I_{\mathrm{other}}$. For simplicity, we set $I_{\mathrm{other}} = f I_{\mathrm{own}}$ [12].

A CAC policy determines the state-dependent call admission vector $\mathbf{u}_c(\mathbf{n}) = (\mathbf{u}_{\mathrm{c}}^{\mathrm{n}}(\mathbf{n}), \mathbf{u}_{\mathrm{c}}^{\mathrm{h}}(\mathbf{n}))$. Admission vector $\mathbf{u}_{\mathrm{c}}^{\mathrm{n}}(\mathbf{n}) = (u_{\mathrm{c}_1}^{\mathrm{n}}, \ldots, u_{\mathrm{c}_J}^{\mathrm{n}})$, $u_{\mathrm{c}_j}^{\mathrm{n}} \in [0, 1]$, represents the decision to admit a new call when $u_{\mathrm{c}_j}^{\mathrm{n}} = 1$ or reject it when $u_{\mathrm{c}_j}^{\mathrm{n}} = 0$. Similarly, a handoff call is admitted when $u_{\mathrm{h}_j}^{\mathrm{c}} = 1$ or rejected when $u_{\mathrm{c}_j}^{\mathrm{h}} = 0$. When $\eta_{\mathrm{sys}}(\mathbf{n}, \mathbf{u}_{\mathrm{r}}(\mathbf{n}))$ approaches 1, the system reaches its capacity threshold $\eta_{\max}$ and $\varphi_{\mathrm{sys}}(\mathbf{n}, \mathbf{u}_{\mathrm{r}}(\mathbf{n}))$ approaches infinity. Dynamic pricing control determines the state-dependent admission price scalar $u_{\mathrm{p}}(\mathbf{n}) \in \mathcal{U}_{\mathrm{p}}$, where $\mathcal{U}_{\mathrm{p}}$ is the set of possible values of $u_{\mathrm{p}}(\mathbf{n})$. This pricing strategy is nondiscriminatory in the sense that the same admission price per unit resource-time applies to all new users, regardless of their service. A *new* call will only be admitted if the user has sufficient WTP and the interference threshold satisfies $\eta_{\mathrm{sys}}(\mathbf{n}+\mathbf{e}_j, \mathbf{u}_{\mathrm{r}}(\mathbf{n}+\mathbf{e}_j)) \leq \eta_{\max} < 1$.

A. Load-Based Pricing

In load-based pricing, users are charged according to their individual load factor (ILF) η_j, which is the ratio of their individual load with respect to the system loading. Assuming that the transmit power of each mobile station (MS) is perfectly controlled based on the receiving level at the Node B, the minimum power that the ith user of service j must transmit in order to achieve (1) is given by $P_j = \eta_j(u_{\mathrm{r}j}(\mathbf{n}))I_{\mathrm{total}}$, where $\eta_j(u_{\mathrm{r}j}(\mathbf{n}))$ is defined as

$$\eta_j\left(u_{\mathrm{r}j}(\mathbf{n})\right) = (1+f)\left(1 + \frac{W}{\left(\frac{E_b}{N_o}\right)_j u_{\mathrm{r}j}(\mathbf{n})\nu_j}\right)^{-1}. \qquad (2)$$

The system load factor is defined as the sum of all individual load factors. With $\mathbf{n}$ users, the system load factor is given by

$$\eta_{\mathrm{sys}}\left(\mathbf{n}, \mathbf{u}_{\mathrm{r}}(\mathbf{n})\right) = \sum_{j=1}^{J} \eta_j\left(u_{\mathrm{r}j}(\mathbf{n})\right) n_j. \qquad (3)$$

When a user requests a call connection of service j with price per ILF-time $u_{\mathrm{p}}(\mathbf{n})$, they will decide to either make a connection request if their budget is sufficient to cover the expected call cost of length $1/\mu_j$ or defer the request otherwise. We denote the probability of having the sufficient WTP as the access probability, i.e., $\alpha_{\mathrm{p}_j}(\mathbf{n}, \mathbf{u}) = \Pr(\Psi_j \geq (u_{\mathrm{p}}(\mathbf{n})\eta_j(u_{\mathrm{r}j}(\mathbf{n}+\mathbf{e}_j)))/\mu_j)$, which can be seen as an arrival gate that controls the flow of price-affected arrivals to the system. λ_j^{n} is the maximum *new* arrival rate, limited only by α_{Pj}. Since handoff calls are preadmitted at another price, their arrival rate will be independent of the current admission price and should never be dropped on the basis of insufficient budget.

B. Interference-Based Pricing

In interference-based pricing, users are charged according to the interference generated by their call. The total interference on the uplink can be estimated using the system load factor defined in (3). The system noise rise can be expressed as

$$\varphi_{\mathrm{sys}}\left(\mathbf{n}, \mathbf{u}_{\mathrm{r}}(\mathbf{n})\right) = -10\log_{10}\left(1 - \eta_{\mathrm{sys}}\left(\mathbf{n}, \mathbf{u}_{\mathrm{r}}(\mathbf{n})\right)\right) \qquad (4)$$

using $I_{\mathrm{total}} = P_{\mathrm{N}} + I_{\mathrm{own}} + I_{\mathrm{other}} = P_{\mathrm{N}} + \sum_{j=1}^{J} P_j$. When the system is empty, the system load factor and noise rise are $\eta_{\mathrm{sys}}(\mathbf{n}, \mathbf{u}_{\mathrm{r}}(\mathbf{n})) = 0$ and $\varphi_{\mathrm{sys}}(\mathbf{n}, \mathbf{u}_{\mathrm{r}}(\mathbf{n})) = 0$ dB, respectively. The system noise rise, defined as the ratio of the total received wideband power to the background thermal noise, is a metric for measuring the total interference in the cell. From (4), the system noise rise increases logarithmically with the system load factor, which depends on the individual load factor of all users. We now propose a metric to measure the amount of interference generated by a call.

Definition 1: The noise rise factor (NRF) $\varphi_j(\mathbf{n}, \mathbf{u}_{\mathrm{r}}(\mathbf{n}))$ of a call with load factor $\eta_j(u_{\mathrm{r}j}(\mathbf{n}))$ is defined as

$$\varphi_j\left(\mathbf{n}, \mathbf{u}_{\mathrm{r}}(\mathbf{n})\right) = \frac{\varphi_{\mathrm{sys}}\left(\mathbf{n}, \mathbf{u}_{\mathrm{r}}(\mathbf{n})\right)}{\eta_{\mathrm{sys}}\left(\mathbf{n}, \mathbf{u}_{\mathrm{r}}(\mathbf{n})\right)}\eta_j\left(u_{\mathrm{r}j}(\mathbf{n})\right) \qquad (5)$$

where $\eta_{\mathrm{sys}}(\mathbf{n}, \mathbf{u}_{\mathrm{r}}(\mathbf{n}))$ and $\varphi_{\mathrm{sys}}(\mathbf{n}, \mathbf{u}_{\mathrm{r}}(\mathbf{n}))$ are the system load factor and noise rise, respectively. Note that the first component gives the noise rise per load factor. Multiplying this component with the load factor of a call gives its individual noise rise.

Authorized licensed use limited to: IEEE Xplore. Downloaded on October 23, 2008 at 20:27 from IEEE Xplore. Restrictions apply.

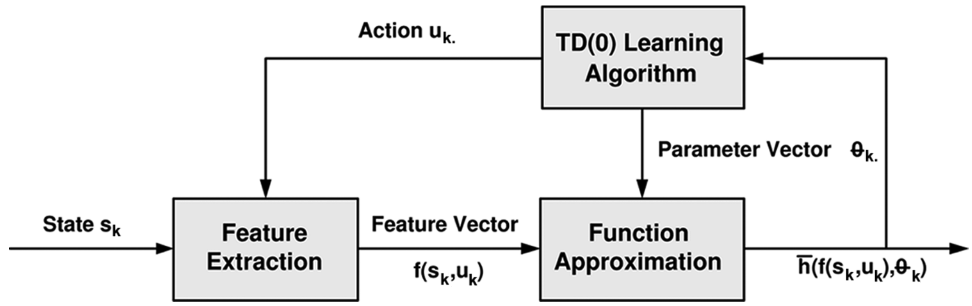

Fig. 1. Modified approximation architecture with an action-dependent feature vector $\boldsymbol{\theta}_k$.

Similar to load-based pricing, when users request a call connection of service j with price $u_{\rm p}(\mathbf{n})$, they will decide to either make a connection request if their budget is sufficient to cover the expected call cost of length $1/\mu_j$, or defer the request otherwise. The access probability of users is given by $\alpha_{{\rm p}_j}(\mathbf{n},\mathbf{u}) = \Pr(\Psi_j \geq (u_{\rm p}(\mathbf{n})\varphi_j(\mathbf{n}+\mathbf{e}_j,\mathbf{u}_{\rm r}(\mathbf{n}+\mathbf{e}_j)))/\mu_j)$. The use of noise rise factor as a basis for charging users in an interference-limited network is motivated by the nonlinear increase in the amount of interference generated by a call as the network approaches its maximum system loading. Specifically, as the system load factor increases, the noise rise factor of a call increases nonlinearly at a positive rate of $10\eta_j(u_{{\rm r}j}(\mathbf{n}))/\eta_{\rm sys}(\mathbf{n},\mathbf{u}_{\rm r}(\mathbf{n}))[(1/(\log 10)(1-\eta_{\rm sys}(\mathbf{n},\mathbf{u}_{\rm r}(\mathbf{n}))))+(\log_{10}(1-\eta_{\rm sys}(\mathbf{n},\mathbf{u}_{\rm r}(\mathbf{n})))/\eta_{\rm sys}(\mathbf{n},\mathbf{u}_{\rm r}(\mathbf{n})))]$ when the parameters of $\eta_j(u_{{\rm r}j})$ of all existing calls remain constant. By contrast, the contribution of one's load factor to the system load factor remains the same due to the additive nature of (3). Therefore, the noise rise factor is a more accurate representation of the users' resource consumption than their individual load factor in networks with soft capacity. Using interference-based pricing, prices increase at the rate of which interference increases. Due to the complexity of the model, we will verify the advantage of interference-based pricing over load-based pricing in Section V.

III. Dynamic Programming Formulation

Any state transition is caused by one of the following events: an arrival of a new call; an arrival of a handoff call; and departure or handoff of an ongoing call. Since we do not keep track of the number of users in other cells, departure and handoff of ongoing calls can be treated as the same event. Let $\Omega = \{\omega | \omega \in \{\mathbf{0}, \omega_j^{\rm n}, \omega_j^{\rm h}, \omega_j^{\rm d}\}, j \in [1, J]\}$ denote the set of possible events, where $\mathbf{0}$, $\omega_j^{\rm n}$, $\omega_j^{\rm h}$, and $\omega_j^{\rm d}$ represent no state transition, a new or handoff call arrival and a departure, respectively. Pricing, resource allocation and CAC are triggered when there is a new or handoff call request. Let $\mathcal{U}_{\rm c}$, $\mathcal{U}_{\rm r}$ and $\mathcal{U}_{\rm p}$ be the set of possible call admission, resource allocation and dynamic pricing actions. When an event $\omega \in \omega_j^{\rm n}, \omega_j^{\rm h}$ occurs, $\mathcal{U}(\mathbf{n},\omega) = \{\mathcal{U}_{\rm c} \times \mathcal{U}_{\rm r} \times \mathcal{U}_{\rm p}\}$ is the set of available actions in state $s = \mathbf{n}$. Using uniformization [17], the continuous-time Markov decision problem (MDP) can be transformed into its discrete-time equivalence with the so-called uniform transition rate, where the total transition rate out of any state is bounded by Λ. The transition probabilities are then given by

$$p(\mathbf{n},\omega,\mathbf{u}) = \begin{cases} u_{{\rm c}_j}^{\rm n}\alpha_{{\rm p}_j}(\mathbf{n},\mathbf{u})\lambda_j^{\rm n}/\Lambda, & \text{if } \omega = \omega_j^{\rm n} \\ u_{{\rm c}_j}^{\rm h}\lambda_j^{\rm h}/\Lambda, & \text{if } \omega = \omega_j^{\rm h} \\ n_j(\mu_j+\nu_j)/\Lambda, & \text{if } \omega = \omega_j^{\rm d} \\ 1-\tau(\mathbf{n})/\Lambda, & \text{otherwise} \end{cases} \tag{6}$$

where $\sum_j u_{{\rm c}_j}^{\rm n}(\alpha_{{\rm p}_j}(\mathbf{n},\mathbf{u})\lambda_j^{\rm n} + \lambda_j^{\rm h}) + n_j(\mu_j+\nu_j) \leq \Lambda$ is the total transition rate out of state $s = \mathbf{n}$. When the system is in state $s = \mathbf{n}$ and an event $\omega \in \Omega$ occurs, a control action $u \in \mathcal{U}(\mathbf{n},\omega)$ is selected. The next state s' is given by the following state transition function:

$$s' = y(\mathbf{n},\omega,\mathbf{u}) = \begin{cases} \mathbf{n}+\mathbf{e}_j, & \text{if } \omega = \omega_j^{\rm n}, u_{{\rm c}_j}^{\rm n} = 1 \\ \mathbf{n}+\mathbf{e}_j, & \text{if } \omega = \omega_j^{\rm h}, u_{{\rm c}_j}^{\rm h} = 1 \\ \mathbf{n}-\mathbf{e}_j, & \text{if } \omega = \omega_j^{\rm d}, \\ \mathbf{n}, & \text{otherwise.} \end{cases} \tag{7}$$

Since premature termination of ongoing calls is usually more undesirable than rejection of new call requests, it has been widely accepted that a system should allocate a higher priority to handoff call requests compared to new call requests. We introduce the term satisfaction revenue (SR) to denote the monetary measure of users' satisfaction with the continuation of a call when a handoff is successful. However, satisfaction revenue is not a real income to the network provider. Let $s' = \mathbf{n}+\mathbf{e}_j$ denote the subsequent state when there a new user is admitted. The immediate revenue collected is

$$g(\mathbf{n},\omega,\mathbf{u}) = \begin{cases} u_{{\rm c}_j}^{\rm n}u_{\rm p}(\mathbf{n})\eta_j \\ \quad \times (u_{{\rm r}j}(s'))/\mu_j, & \text{if } \omega = \omega_j^{\rm n} \text{ (ILF-based)} \\ u_{{\rm c}_j}^{\rm n}u_{\rm p}(\mathbf{n})\varphi_j \\ \quad \times (s',\mathbf{u}_{\rm r}(s'))/\mu_j, & \text{if } \omega = \omega_j^{\rm n} \text{ (NRF-based)} \\ u_{{\rm c}_j}^{\rm h}SR & \text{if } \omega = \omega_j^{\rm h}. \end{cases} \tag{8}$$

In order to reflect the higher importance of accepting a handoff call, SR should be greater than the actual revenue collected when a new call request is accepted. The average reward-to-go function, known as the Bellman's equation, is given by

$$J^* + h(s) = \max_{\mathbf{u}\in\mathcal{U}(s,\omega)} \left[\sum_{\omega\in\Omega} p(s,\omega,\mathbf{u}) \times [g(s,\omega,\mathbf{u}) + h(y(s,\omega,\mathbf{u}))]\right]. \tag{9}$$

J^* and $h(s)$ denote the optimal average reward and the differential reward rate of state s, respectively. A stage here means a transition in the uniformized chain. The optimal expected reward per stage is independent of the initial state. Standard average-reward DP theory applies and there exists a stationary policy which is optimal [17].

IV. NDP Formulation

NDP refers to approximate methods that centre around the evaluation and approximation of the optimal cost-to-go function (9), possibly

Authorized licensed use limited to: IEEE Xplore. Downloaded on October 23, 2008 at 20:27 from IEEE Xplore. Restrictions apply.

through simulation and/or the use of neural networks. Instead of computing the differential reward function $h(s)$ for every state $s \in \mathcal{S}$, NDP uses a compact representation $\tilde{h}(\cdot,\boldsymbol{\theta})$ to approximate $h^*(\cdot)$, using parameter vector $\boldsymbol{\theta}$. Naturally, we want to define the general structure of $\tilde{h}(\cdot,\boldsymbol{\theta})$ and calculate parameter vector $\boldsymbol{\theta}$ so as to minimize the error between the functions $h^*(\cdot)$ and $\tilde{h}(\cdot,\boldsymbol{\theta})$. The process of tuning parameters $\boldsymbol{\theta}$ is often referred as training or learning. The average reward per time J^* is approximated by tunable scalar $\tilde{J}$. If $\tilde{h}(\cdot,\boldsymbol{\theta})$ and $\tilde{J}$ are close enough to the $h^*(s)$ and J^*, then the greedy control policy induced is, in some sense, close to an optimal policy. Hereafter, we denote the k^{th} step estimate of $\tilde{h}(\cdot,\boldsymbol{\theta})$ and $\tilde{J}$ as $\tilde{h}(\cdot,\boldsymbol{\theta}_k)$ and $\tilde{J}_k$, respectively. There are two major parts in NDP: an approximation architecture to define the structure of $\tilde{h}(\cdot,\boldsymbol{\theta}_k)$; and a learning algorithm for tuning $\tilde{h}(\cdot,\boldsymbol{\theta}_k)$ and $\tilde{J}$.

Approximation Architecture: The first task is to select an appropriate approximation architecture for $\tilde{h}(s_k,\theta_k)$, which is a functional form involving a number of free parameters that are tuned to provide the best fit of $h^*(s_k)$. It is often the case that the complexity of function $\tilde{h}(s_k,\theta_k)$ can be reduced by feeding a set of *features* of the state into an approximation architecture. The task of selecting the appropriate set of features is usually problem-dependent. Based on various experiments, we propose a modified version of the general feature-based approximation architecture used in [13] that includes the decision/action $\mathbf{u}_k$ in the feature vector $f(\cdot)$. The modified model is illustrated in Fig. 1. This proposal is based on the properties of the Bellman's equation in (9), where the selection of an action vector not only results in some immediate reward but also affects the reward obtained in future stages. The set of action-dependent features is a mapping of $f_l : \mathcal{S},\mathcal{U} \rightarrow R$, $l = 2,\ldots,L$, defined as

$$f_l(s_k,\mathbf{u}_k) = \begin{cases} u^n_{cj}\alpha_{\mathrm{p}_j}(s_k,\mathbf{u}_k) \\ \quad \times\lambda^{\mathrm{n}}_j u_{\mathrm{p}j}\eta_j(\mathbf{u}_{\mathrm{r}k}), & \text{if } l \in \mathcal{L} \text{ (ILF-based)} \\ u^n_{cj}\alpha_{\mathrm{p}_j}(s_k,\mathbf{u}_k) \\ \quad \times\lambda^{\mathrm{n}}_j u_{\mathrm{p}j}\varphi(s_k,\mathbf{u}_{\mathrm{r}k}), & \text{if } l \in \mathcal{L} \text{ (NRF-based)} \\ \sum_{j=1}^{J} u^h_{cj}\lambda^{\mathrm{h}}_j SR & \text{if } l = L = J+2 \end{cases} \tag{10}$$

where $\mathcal{L} = \{l : 2 \leq l \leq J+1\}$. The first feature is usually set as a scalar offset, i.e., $f_1(s_k,\mathbf{u}_k) = 1$. The next J features are the future reward rates due to new users for every class $j = 1,\ldots,J$. The last feature is the sum of future reward rates due to handoff users. Given a collection of $f(s_k,\mathbf{u}_k) = (f_1(s_k,\mathbf{u}_k),\ldots,f_L(s_k,\mathbf{u}_k))$, we approximate for $\tilde{h}(f(s_k,\mathbf{u}_k),\theta_k)$ instead of $\tilde{h}(s_k,\theta_k)$. We use a linear approximation architecture in the form of $\tilde{h}(f(s_k,\mathbf{u}_k),\boldsymbol{\theta}_k) = \boldsymbol{\theta}_k^T f(s_k,\mathbf{u}_k)$. The dimension of the parameter vector $\boldsymbol{\theta}$ is equal to the number of features L, and the learning problem becomes a linear regression problem.

Learning Algorithm: We will use the TD(0) algorithm for average reward problems [18], [19], which preserves the same convergence properties and error guarantees as its discounted version. The TD(0) algorithm belongs to the class of temporal difference learning algorithms, often known as the $\mathrm{TD}(\lambda)$. At simulation step $k \leq N$, $\tilde{h}(f(s_k,\mathbf{u}_k),\boldsymbol{\theta}_k)$ is used as an approximation of $h^*(s_k)$. Suppose that $\nabla_\theta(h(f(s_k,\mathbf{u}_k),\theta_k))$ exists for every $s_k \in \mathcal{S}$ and $\theta_k \in \mathcal{R}^K$, $\theta_0 \in \mathcal{R}^K$, $J_0 \in \mathcal{R}$, and $s_0 \in \mathcal{S}$, we generate θ_k and J_k using the recursive procedure.

Step 1) Assume that we are given state s_k and parameter vector $\boldsymbol{\theta}_k$, obtain the event $\omega_{k+1} \in \Omega$ according to (6).

Step 2) Choose an action vector $\mathbf{u}_k \in \mathcal{U}(s_k,w_{k+1})$ that satisfies

$$\mathbf{u}_k = \arg\max_{\mathbf{u}_k\in\mathcal{U}}\left[g(s_k,\omega_{k+1},\mathbf{u}) + \tilde{h}\left(f\left(s'_{k+1},\mathbf{u}_k\right),\boldsymbol{\theta}_k\right)\right] \tag{11}$$

using $s'_{k+1} = y(s_k,\omega_{k+1},\mathbf{u}_k)$. Each potential decision $\mathbf{u}_k$ is evaluated in the process of feature extraction.

Step 3) Set $s_{k+1} = y(s_k,\omega_{k+1},\mathbf{u}_k)$, $d_k = g(s_k,\omega_{k+1},\mathbf{u}_k) - (t_{k+1} - t_k)\tilde{J}_k + \tilde{h}(f(s_{k+1},\mathbf{u}_k),\boldsymbol{\theta}_k) - \tilde{h}(f(s_k,\mathbf{u}_k),\boldsymbol{\theta}_k)$. Update vector $\boldsymbol{\theta}_k$ to $\boldsymbol{\theta}_{k+1} = \boldsymbol{\theta}_k + \gamma_k d_k \nabla_{\boldsymbol{\theta}}(\tilde{h}(f(s_k,\mathbf{u}_k),\boldsymbol{\theta}_k))$ and scalar $\tilde{J}_k$ to $\tilde{J}_{k+1} = \tilde{J}_k + \tau_k(g(s_k,\omega_{k+1},\mathbf{u}_k) - (t_{k+1} - t_k)\tilde{J}_k)$.

Step 4) Return to step 1 if $k \leq N$.

Scalar d_k is known as the temporal difference corresponding to the transition from s_k to s_{k+1}. The terms γ_k and τ_k are small step size parameters. Under a fixed policy and standard diminishing step size conditions, J_k and $\boldsymbol{\theta}_k$ will converge to the average reward J^* and vector $\boldsymbol{\theta}$. The algorithm presented is known as optimistic TD(0) because the parameter vector $\boldsymbol{\theta}_k$ is updated according to the greedy action chosen in (11) during each step of the simulation. This algorithm has been widely used in practice, albeit its convergence properties have never been studied thoroughly [13], [19].

V. Numerical Results and Discussions

In this section, we compare the following policies: Static ILF-based Pricing (S-ILF), Always Accept and Average Transmission Rates; Static NRF-based Pricing, Always Accept and Average Transmission Rates (S-NRF); Optimal ILF-based Pricing, CAC and Rates (O-ILF); and Optimal NRF-based Pricing, CAC, and Rates (O-NRF). We simulate a system of three services: two AMR voice services with different range of operating transmission rates and a data service. The AMR speech coder has eight source rates: 12.2, 10.2, 7.95, 7.40, 6.70, 5.90, 5.15, and 4.75 kbps. Similar to [12], the system parameter values used are $W = 3.84$ Mcps; $f = 0.55$; $\mathbf{E_b/N_0} = (5.0, 5.0, 1.5)$ dB; and $\boldsymbol{\nu} = (0.67, 0.67, 1.00)$. Other parameters are $SR = 50$ per handoff call; $\eta_{\max} = 0.98$; $\mathcal{R}^{\min} = (7.40, 4.75, 16.0)$ kbps; $\mathcal{R}^{\max} = (12.2, 12.2, 64.0)$ kbps; $\lambda^{\mathrm{n}} = (5, 10, 10)$; $\lambda^{\mathrm{h}} = (1, 2, 2)$; and $\boldsymbol{\mu} + \boldsymbol{\gamma} = (5, 5, 3)$. The choice of step sizes η and γ are crucial to convergence and after some trial and error, they are set to 10^{-3} and 10^{-8}, respectively, throughout the simulation. For all cases, we use the same random number seed and run the simulation for $N = 1.5 \times 10^6$ steps. Although we do not need to run the TD(0) algorithm for static policies S-ILF and S-NRF, their average reward can be approximated using the update rule for J_k in step 3 of the algorithm.

To ensure a fair comparison among the policies mentioned, the same set of WTP per unit time is used for all simulations. The mean WTP per unit time of each service is proportional to its average transmission rate. We assume that the WTP of users can be fitted into a Weibull distribution with parameters shape β_j and scale ζ_j using mean WTP Ψ_j. This distribution is chosen because of its versatility to take up the characteristics of other types of distributions. The Weibull distribution is one of the most widely used lifetime distributions in reliability engineering [20]. Within the telecommunications framework, it has been used to model the traffic characteristics of packet audio streams in [21] and to simulate data traffic in [22].

Using the Weibull distribution, we can deduct the range of prices per ILF-time or NRF-time using $\alpha_{\mathrm{p}}^{\min} \leq \alpha_{\mathrm{p}_j}(s_k,\mathbf{u}_k) \leq \alpha_{\mathrm{p}}^{\max}$, where $\alpha_{\mathrm{p}}^{\min}$ and $\alpha_{\mathrm{p}}^{\max}$ are the minimum and maximum access probabilities set by the operator. Then, $u_{\mathrm{p}}(\mathbf{n})$ can be calculated using access probability α_{p_j}. The minimum price that corresponds to $\alpha_{\mathrm{p}}^{\max}$ should be set such as to recover the cost needed deliver the service. We set $\alpha_{\mathrm{p}}^{\min}$ and $\alpha_{\mathrm{p}}^{max}$ to 0.1 and 0.8, respectively, and select 35 uniformly distributed prices between them as the price decision space, $\mathcal{U}_{\mathrm{p}}$. The chosen size, denoted as $\#\mathcal{U}_{\mathrm{p}}$, is based on various experiments that indicate that further increase of the pricing space will not provide significantly better results.

The convergence of average reward per unit time $\tilde{J}_k$ and parameter vector $\boldsymbol{\theta}_k$ for O-NRF during training are shown in Fig. 2. The average reward $\tilde{J}_N$ and proportion of reward obtained are illustrated

Authorized licensed use limited to: IEEE Xplore. Downloaded on October 23, 2008 at 20:27 from IEEE Xplore. Restrictions apply.

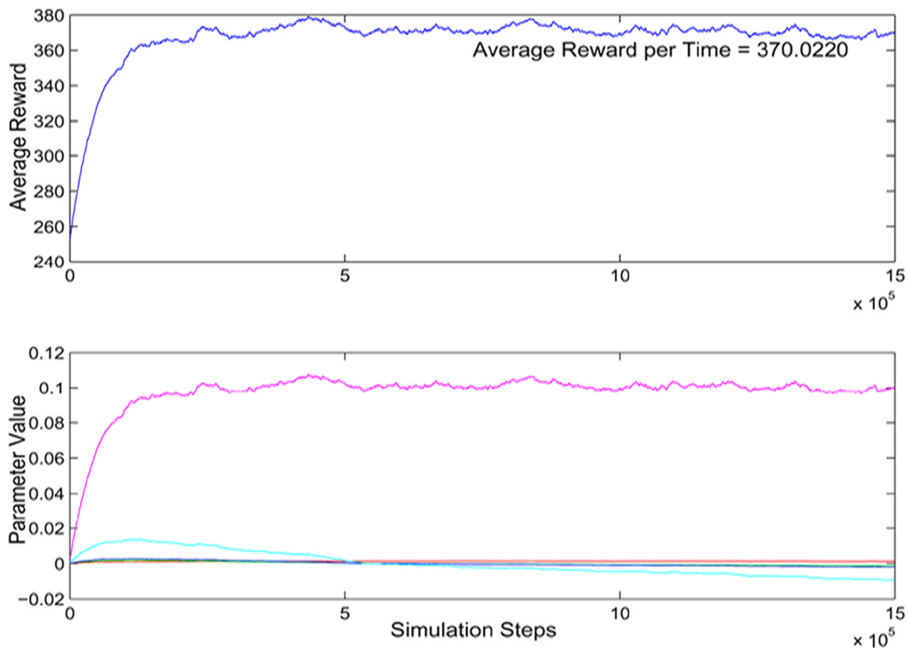

Fig. 2. Average reward per time $\tilde{J}_k$ and $\boldsymbol{\theta}_k$ parameter values, averaged for every 250 steps, for O-NRF.

in Fig. 3(a). The latter is the ratio between the average reward and its potential reward if no users are blocked or dropped due to insufficient budget or the violation of interference threshold. Optimal policy O-NRF accumulated the highest average reward and proportion of reward obtained, followed by O-ILF, S-NRF and S-ILF. The improvement of O-NRF is about 38% over S-ILF, 27% over S-NRF, and 19% over O-ILF. Unlike ILF [see Fig. 3(b)], which remains constant regardless of the level of interference in a network, NRF increases exponentially as the system load factor approaches 1 and noise rise approaches infinity accordingly. Even when static price per unit NRF is used, the price per unit time will rise with the level of system interference because the noise rise factor of a call has increased. This helps to prevent low-WTP users from entering the system when the interference level is high and further aggravating the situation.

The access, blocking and dropping probabilities of all policies are given in Fig. 3(c). The average access probability is lower using NRF-based pricing compared to that of ILF. This reemphasises the earlier point about the exponential increase of price during high interference level in the network. The higher access probability when O-NRF is used, compared to S-NRF, is due to the flexibility of the optimal policy to offer low prices to users when interference is low in the network. Even though the average access probabilities of O-NRF and S-ILF are close, the optimal policy produces a significantly higher average reward due to optimal resource allocation and CAC, in addition to interference-based pricing. The NRF-based policies are also far more effective in controlling the blocking of new users and dropping of handoff users. The blocking and dropping probabilities decrease dramatically to negligible values when interference-based pricing is used in S-NRF and O-NRF. The results from Fig. 3(d) affirm the interference-based pricing as an effective mechanism for congestion control. The load factor of all policies is well below its constraint $\eta_{\max} = 0.98$, which translates to a maximum noise rise of about 17 dB. However, the system load factor and noise rise of O-NRF and S-NRF are notably lower. The flexibility to adjust price according to the state of the network under O-NRF provides more benefits. Under O-NRF, system resources are used more effectively by allowing more users access to service, which is evident through the higher average load factor and noise rise obtained compared to S-NRF.

VI. Conclusion and Discussions

We have formulated an integrated dynamic pricing and RRM problem in an interference-limited network as a NDP problem. A new parameter noise rise factor is suggested as a basis for setting price in an interference-limited network. Unlike existing pricing schemes that charge users based on their transmission rate or load factor, this new parameter can effectively capture the positive, nonlinear relationship between one's resource usage and the interference generated as the system loading increases. We have verified that the interference-based pricing policy improves average reward and blocking and dropping probabilities over load-based pricing schemes, even when static pricing is used. Our proposal for a dynamic admission pricing scheme also minimizes the additional accounting and billing overheads often associated with dynamic pricing schemes because the system only needs to keep track of users' admission price, which will be honoured throughout their call.

We also modified the conventional feature-based approximation architecture to have the effects of each decision on future reward rate evaluated before the decision is made; and successfully adapted the average-reward TD(0) algorithm to a pricing problem. The problem considered involves an offline computation of the average reward and parameter vector with some knowledge of parameters like the arrival rates of the users. Once the approximation obtained offline via learning is satisfactory, it can be used to generate decisions fast enough for use in real time. To adapt this for an online implementation, $\boldsymbol{\lambda}^n$ and $\boldsymbol{\lambda}^h$ have to be estimated using parameter estimation techniques such as maximum likelihood estimation. The estimated parameters $\hat{\boldsymbol{\lambda}}^n$ and $\hat{\boldsymbol{\lambda}}^h$ will then be fed to the feature extraction module. Fig. 4 shows the suggested approach for an online adaptation of this problem. In this correspondnece, we have only considered linear features in terms of the future revenue rate of the system. The extraction of nonlinear features may provide better approximation and can be considered in future work.

Authorized licensed use limited to: IEEE Xplore. Downloaded on October 23, 2008 at 20:27 from IEEE Xplore. Restrictions apply.

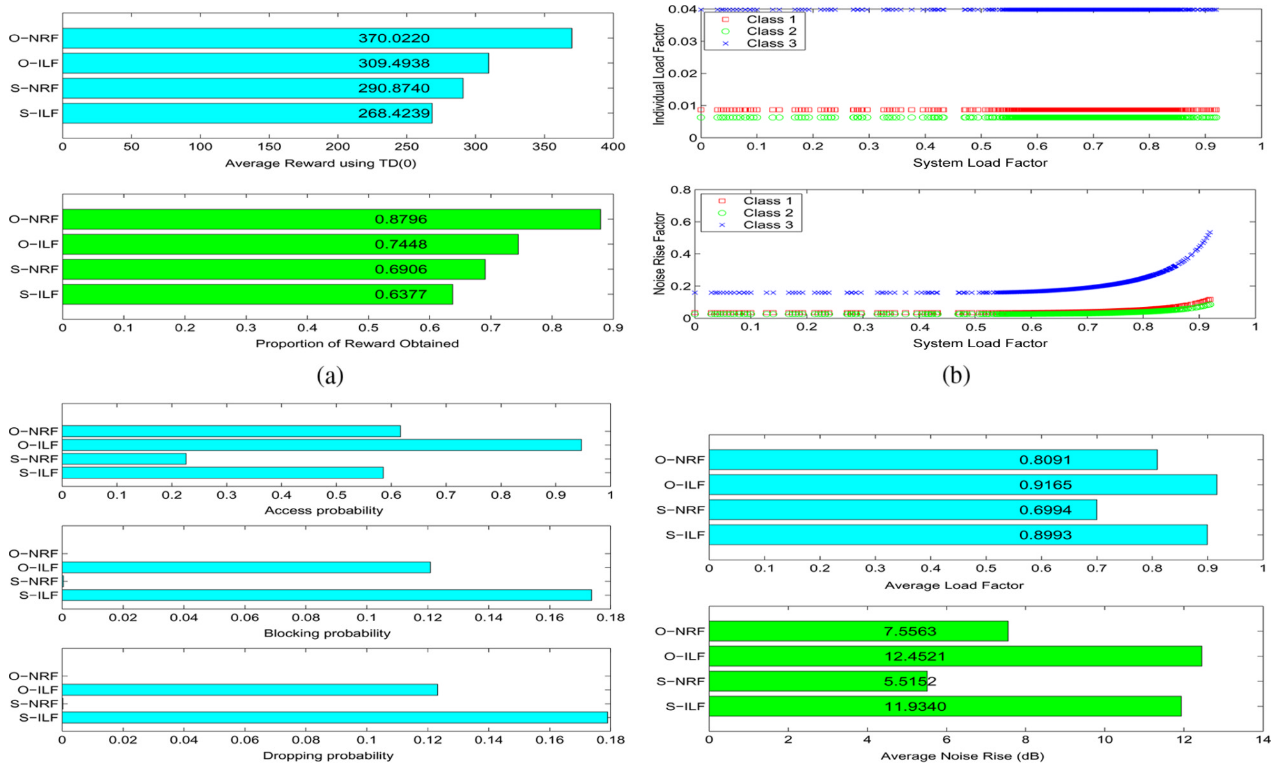

Fig. 3. Simulation results. (a) Average reward and proportion of reward obtained. (b) The relationship between ILF/NRF and system load factor. (c) Blocking, dropping, and access probabilities. (d) Average load factor and noise rise.

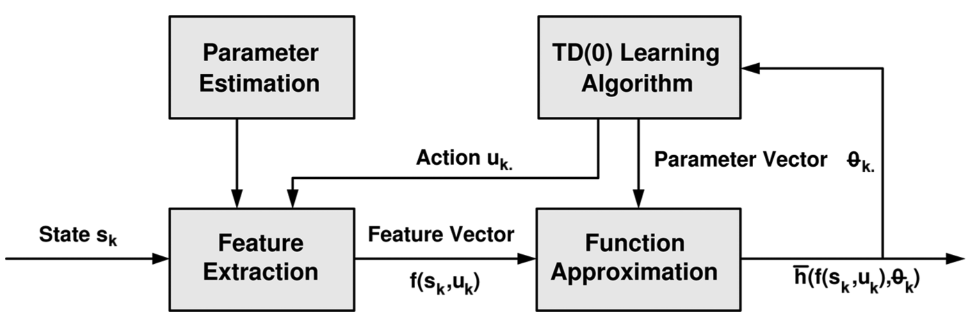

Fig. 4. Modified feature-based approximation architecture with parameter estimation.

This interference-based pricing and RRM policy $\mathbf{u}$ can then be implemented in practical WCDMA systems using the following steps. Suppose the system is in state $s = \mathbf{n}$. Based on the pricing policy $\mathbf{u}_{\mathrm{p}}(s)$, the system announces a set of service-dependent admission prices to the users in the cell. The price per unit time of service j is given as $\mathbf{u}_{\mathrm{p}}(s)\varphi_j(s+\mathbf{e}_j, \mathbf{u}_{\mathrm{r}}(s+\mathbf{e}_j))$, where the noise rise factor of service j, $\varphi_j(s+\mathbf{e}_j, \mathbf{u}_{\mathrm{r}}(s+\mathbf{e}_j))$, can be computed according to the load estimation techniques based on the wideband received power in [12]. When a new request for service arrives, it will be admitted with the price only if $u_{\mathrm{rj}}(s) = 1$. For handoff calls, their assigned price per unit time during their earlier admission will be maintained throughout the call. When a call is admitted, the state changes to $s' = s+\mathbf{e}_j$ and the transmission rates of all users are given by vector $\mathbf{u}_{\mathrm{r}}(s')$. Finally, our pricing policy relies on the operator's ability to estimate the statistical distribution of users' WTP. In reality, this information can be obtained in a number of ways. For example, the budget information can be extracted from the operator's historical data on users' spending patterns. By offering suitable incentives to users, they can also be encourage to share their WTP. The cost-to-go approximation will then continue to improve as the system operates in real time.

References

[1] J. Mackie-Mason and H. Varian, "Pricing congestible network resources," *IEEE J. Sel. Areas Commun.*, vol. 13, pp. 1141–1148, Sep. 1995.

[2] R. Gibbens and F. Kelly, "Resource pricing and the evolution of congestion control," *Automatica*, vol. 35, pp. 1969–1985, 1999.

[3] I. Paschalidis and J. Tsitsiklis, "Congestion-dependent pricing of network services," *IEEE/ACM Trans. Netw.*, vol. 8, pp. 171–184, Apr. 2000.

[4] I. Paschalidis and Y. Liu, "Pricing in multiservice loss networks: Static pricing, asymptotic optimality, and demand substitution effects," *IEEE/ACM Trans. Netw.*, vol. 10, pp. 425–437, Jun. 2002.

[5] L. DaSilva, "Pricing for QoS-enabled networks: A survey," *IEEE Commun. Surv.*, pp. 2–8, Second Quarter 2000.

[6] C. Saraydar, N. Mandayam, and D. Goodman, "Efficient power control via pricing in wireless data networks," *IEEE Trans. Commun.*, vol. 50, pp. 291–303, Feb. 2002.

Authorized licensed use limited to: IEEE Xplore. Downloaded on October 23, 2008 at 20:27 from IEEE Xplore. Restrictions apply.

[7] C. Saraydar, N. Mandayam, and D. Goodman, "Pricing and power control in a multicell wireless data network," *IEEE J. Sel. Areas Commun.*, vol. 19, pp. 1883–1892, Oct. 2001.
[8] M. Xiao, N. Shroff, and E. K. Chong, "A utility-based power-control scheme in wireless cellular systems," *IEEE/ACM Trans. Netw.*, vol. 11, pp. 210–221, Apr. 2003.
[9] J. W. Lee, R. Mazumdar, and N. Shroff, "Downlink power allocation for multi-class CDMA wireless networks," *IEEE/ACM Trans. Netw.*, vol. 13, pp. 854–867, Aug. 2005.
[10] P. Liu, P. Zhang, S. Jordan, and M. L. Honig, "Single-cell forward link power allocation using pricing in wireless networks," *IEEE/ACM Trans. Wireless Commun.*, vol. 3, pp. 533–543, Mar. 2004.
[11] C. Zhou, P. Zhang, M. L. Honig, and S. Jordan, "Two-cell power allocation for downlink CDMA," *IEEE Trans. Wireless Commun.*, vol. 3, Nov. 2004.
[12] H. Holma and A. Toskala, Eds., *WCDMA for UMTS*. New York: Wiley, 2001.
[13] D. Bertsekas and J. Tsitsiklis, *Neuro-Dynamic Programming*. Boston, MA: Athena Scientific, 1996.
[14] P. Marbach, O. Mihatsch, and J. N. Tsitsiklis, "Call admission control and routing in integrated service networks using neuro-dynamic programming," *IEEE J. Sel. Areas Commun.*, vol. 18, pp. 197–208, Feb. 2000.
[15] B. Van Roy, D. Bertsekas, Y. Lee, and J. N. Tsitsiklis, "A neuro-dynamic programming approach to retailer inventory management," in *Proc. 36th IEEE Conf. Decision Contr.*, 1997, pp. 4052–4057.
[16] N. Hedge and E. Altman, "Capacity of multiservice WCDMA networks with variable GoS," in *IEEE Wireless Commun. Netw. Conf. 2003*, Mar. 16–20, 2003, vol. 2, pp. 1402–1407.
[17] D. P. Bertsekas, *Dynamic Programming and Optimal Control*. Boston, MA: Athena Scientific, 1995, vol. 1 and 2.
[18] R. Sutton, "Learning to predict by the methods of temporal differences," *Mach. Learn.*, vol. 3, pp. 9–44, 1988.
[19] J. Tsitsiklis and B. Van Roy, "Average cost temporal-difference learning," in *Proc. 36th Conf. Decision Contr.*, Dec. 1997, pp. 498–502.
[20] A. Papoulis and S. U. Pillai, *Probability, Random Variables and Stochastic Processes*, 4th ed. New York: McGraw-Hill, 2002.
[21] C. Chuah and R. Katz, "Characterizing packet audio streams from internet multimedia applications," in *IEEE Int. Conf. Commun.*, Apr. 2002, pp. 1199–1203.
[22] A. Silva and G. Mateus, "Performance analysis for data service in third generation mobile telecommunication networks," in *Proc. 35th Ann. Simulation Symp.*, Apr. 2002, pp. 227–234.

Stability Analysis and Design of Impulsive Control Systems With Time Delay

Zhichun Yang and Daoyi Xu

***Abstract*—A class of impulsive control systems with time-varying delays is considered. By establishing an impulsive delay differential inequality, we analyze the global exponential stability of the impulsive delay systems and estimate the exponential convergence rate. On the basis of the analysis, a design procedure of impulsive controller is presented. The designed impulsive controller not only can globally exponentially stabilize the time delay systems, but also can control the exponential convergence rate of the systems. Two numerical examples are given to illustrate the effectiveness of the method.**

***Index Terms*—Exponential stability, impulsive control, impulsive delay differential inequalities, time delay.**

I. Introduction

Stabilization problem of linear or nonlinear dynamic systems with time delay is receiving much attention. The existence of a delay in a nonlinear system may induce more complex dynamical behaviors such as instability, oscillations, and chaos, e.g., in Mackey-Glass model [1] and Chua's circuit system with delay [2]. Some important control methods have been developed for stabilizing dynamic systems without delay or with delay, which include state feedback control [3], [4], adaptive control [5], fuzzy control [6], variable structure control [7], etc.

Recently, impulsive control method has attracted increasing interests in engineering, economics, medicine, and biology. The examples include ecosystems management [8], orbital transfer of satellite [9], optimal control of economic systems [10], synchronization of chaos-based secure communication systems [11], [12], and so on. The main idea of impulsive control is to change the states of continuous dynamic systems via discontinuous control inputs at certain time moments, which is actually a scheme of hybrid control (see also, [13]–[15]). In some cases, the scheme of hybrid impulsive control may be more efficient than one of continuous control. In chaotic secure communication systems, for instance, impulsive control is attractive since it allows the stabilization and synchronization of a chaotic system using only small control impulses [11], [12]. According to [16], the efficiency of the bandwidth usage is improved more greatly by using impulsive control than by continuous control for the synchronization in the chaos-based secure communication systems. On the other hand, time delays frequently appear in various dynamical systems, for instance, delay effects are inevitable in the mentioned communication systems due to the finite switching speed of the hardware. This motivates the present investigation of impulsive control problem for time delay systems, which is generally governed by the stability theory of impulsive differential equations. Furthermore, since absolute

Manuscript received March 19, 2005; revised April 6, 2006 and February 5, 2007. Recommended by Associate Editor M. Demetriou. This work was supported in part by the NSFC by Grant 10671133, and by the Scientific Research Fund of Sichuan Provincial Education Department by Grant 2006B071.

Z. Yang is with the Key Lab of Operational Research and Systems Engineering, Chongqing Normal University, Chongqing 400047, China and also with the Chinese Academy of Sciences, Beijing 100080, China (e-mail: zhichy@yahoo.com.cn).

D. Xu is with the Yangtze Center of Mathematics, Sichuan University, Chengdu 610064, China (e-mail: daoyixucn@yahoo.com).

Color versions of one or more of the figures in this note are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TAC.2007.902748

0018-9286/$25.00 © 2007 IEEE

Authorized licensed use limited to: IEEE Xplore. Downloaded on October 23, 2008 at 20:27 from IEEE Xplore. Restrictions apply.