Computer Communications 32 (2009) 317–323

Contents lists available at ScienceDirect

Computer Communications

journal homepage: www.elsevier.com/locate/comcom

Utilizing call admission control for pricing optimization of multiple service classes in wireless cellular networks

Okan Yilmaz *, Ing-Ray Chen

Computer Science Department, Virginia Tech, 7054 Haycock Road, Falls Church, VA 22043, United States

A R T I C L E I N F O

Article history:
Received 29 January 2007
Received in revised form 30 October 2008
Accepted 2 November 2008
Available online 9 November 2008

Keywords:
Pricing optimization
Call admission control
Wireless networks
Quality of service
Revenue optimization
Service classes

A B S T R A C T

In this paper, we utilize admission control algorithms designed for revenue optimization with QoS guarantees to derive optimal pricing of multiple service classes in wireless cellular networks. A service provider typically adjusts pricing only periodically. Once a "global" optimal pricing is derived, it would stay static for a period of time, allowing users to be charged with the same rate while roaming. We utilize a hybrid partitioning-threshold admission control algorithm to analyze a pricing scheme that correlates service demand with pricing, and to periodically determine optimal pricing under which the system revenue is maximized while guaranteeing that QoS requirements of multiple service classes are satisfied.

1. Introduction

Next generation wireless networks will provide diverse multimedia services to mobile users, including real-time services such as video and audio streaming, and non-real-time services such as image and text access. As these multiple multimedia service classes have distinct quality of service (QoS) requirements, it is important to design admission control algorithms that admit roaming users of different service types to satisfy their distinct QoS requirements while maximizing the utilization of system resources.

The blocking probability of new calls and the dropping probability of handoff calls are two important QoS metrics. Mobile users in a cellular network establish a connection through their local base station. A base station has a fixed number of wireless channels and can only support a limited number of connections (or calls) simultaneously due to bandwidth limitations. A handoff occurs when a mobile user with an ongoing connection leaves the current cell and enters another cell. Thus, an ongoing, incoming connection may be dropped during a handoff if there is insufficient bandwidth in the new cell to support it. We can reduce the handoff call drop probability by rejecting new connection requests. However, this increases the new call blocking probability. Thus, there is a tradeoff between the handoff and new call blocking probabilities.

Call admission control for single-class network traffic, such as voice, has been studied extensively [6–8,14]. For multiple service classes, call admission algorithms offered in [5,13,16,17] make acceptance decisions for new and handoff calls to satisfy QoS requirements in order to keep the dropping probability of handoff calls and the blocking probability of new calls below a specified threshold. All these algorithms concern QoS requirements, not pricing or revenue issues of service classes. Chen et al. [3] first proposed the concept of maximizing the "payoff" of the system through admission control in the context of multimedia services. Recently, Chen et al. [4] developed a class of admission control algorithms integrated with pricing with QoS guarantees based on partitioning [3] and threshold-based [13] and a hybrid partitioning-threshold algorithm combining both for revenue optimization in wireless networks. However, these studies aimed at allocating system resources to maximize the revenue received given that a fixed price has been assigned by the service provider. In this paper, we address the issue of determining optimal pricing. We utilize admission control algorithms to derive optimal pricing of multiple service classes in mobile wireless networks for revenue optimization with QoS guarantees.

The area of optimal pricing for service multiple classes in wireless networks is relatively unexplored. Hou et al. [9] proposed a dynamic pricing approach in response to changing call arrival rates to satisfy QoS requirements. We dispose dynamic pricing as a valid approach since changing pricing during call services is disturbing to callers. Instead, we consider pricing is static for each service class and is changed only periodically. Aldebert et al. [1] presented an empirical study that reveals the relationship between pricing and demand for residential telecommunication service. Rappoport

* Corresponding author. Tel.: +1 571 434 5685.
E-mail addresses: oyilmaz@vt.edu (O. Yilmaz), irchen@vt.edu (I.-R. Chen).

0140-3664/$ - see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.comcom.2008.11.001

et al. [15] analyzed a consumer survey to estimate the demand for wireless internet access. Keon and Anandalingam [10] proposed an optimal pricing approach in the context of wired networks considering a pricing scheme that would charge customers for the use of a connection through a sequence of switches based on a pricing-demand relationship identified in [1,12] to optimize revenue. Our paper concerns pricing for service-oriented classes charged by the amount of time a call uses the service, rather than being charged per connection.

The basic idea for finding optimal pricing is that for each cell we periodically and statistically determine a set of reference parameter values for each service class, including the arrival/departure rates of new calls, and the arrival/departure rates of handoff calls, characterizing the cell's current operating and workload conditions. Then, by utilizing a pricing-demand relation function as suggested in [11,12], we adjust pricing relative to the reference pricing, thereby predicting the arrival/departure rates of each service class relative to the service class's reference arrival/departure rates in each cell. For each price adjustment, a call admission control algorithm developed in the paper is utilized by each cell to determine the maximum revenue obtainable. The end product is a table generated periodically by each cell listing all candidate price combinations considered by the service provider, and the associated "local" revenue obtainable as a result of executing an admission control algorithm cognizant of revenue optimization. The tables from individual cells then will be merged by the network service provider to determine "global" optimal pricing across all the cells in the system. The optimal price per class determined this way is the same across cells in the system, so a user would be charged with the same rate for a service as it roams across cells in the system. When the price of a service call is changed, the new rate would be applied to all cells in the system. Thus, a user would not have to deal with the annoying issue associated with dynamic pricing that it may be charged more when it roams into a cell that is overloaded, so a higher price may be imposed by that cell to reduce traffic in the cell.

The goal of this paper is to utilize admission control algorithms designed for revenue optimization with QoS guarantees to derive optimal pricing. A service provider typically adjusts pricing only periodically. Once a "global" optimal pricing is derived, it would stay static for a period of time, allowing users to be charged with the same rate while roaming. We show that a hybrid admission control algorithm combining the benefits of partitioning and threshold-based call admission control would perform the best in terms of pricing optimization to maximize the revenue earned with QoS guarantees to multiple service classes over a wide range of input parameters characterizing the operating conditions.

The rest of the paper is organized as follows. Section 2 states the system model and gives assumptions used in characterizing the operational environment of a wireless network. Section 3 presents our pricing scheme and methodology for finding optimal pricing. Section 4 describes a class of call admission control algorithm designed for revenue optimization with QoS guarantees for mobile wireless networks. Section 5 discusses applicability in terms of how call admission control can be utilized to derive optimal pricing for multiple service classes and presents numerical results with physical interpretation given. Finally, Section 6 summarizes the paper and outlines some future research areas.

2. System model

A cellular network consists of a number of cells, each of which has a base station at the center to provide network services to mobile hosts within the cell. We assume there exists a number of distinct *service classes*, $S^1, S^2, \ldots, S^n$, characterized by service type attribute. A service type can be *real-time* or *non-real-time*. Furthermore, there are two call types: *handoff* and *new*. Of these call types, handoff calls always have higher priority than new calls. Each service type requires a number of bandwidth channels to satisfy its intrinsic bandwidth QoS requirement. Each combination of service type and call type may also impose a system-wide QoS requirement. For example, the handoff call drop probability of real-time service type being less than 2% can be a QoS requirement. Assume that for each service class, say i, a QoS constraint exists on the handoff call blocking probability threshold $B_h^i t$ and the new call blocking probability threshold $B_n^i t$.

From the perspective of a single cell, each service class/call type combination is characterized by its arrival rate, and departure rate. Let λ_n^i denote the arrival rate of *new* calls of service class i and μ_n^i be the corresponding departure rate. Similarly, let λ_h^i denote the arrival rate of *handoff* calls of service class i, and μ_h^i be the corresponding departure rate. These parameters can be determined by inspecting statistics collected by the base station in the cell and by consulting with base stations of neighbor cells [4]. Without loss of generality we assume that a cell has C channels where C can vary depending on the amount of bandwidth available in the cell. When a call of service class i enters a handoff area from a neighboring cell, a handoff call request is generated. Each call has its specific QoS bandwidth requirement dictated by its service traffic type attribute. Assume that a service call of service class i requires k^i channels regardless of its call type.

Another parameter that is associated with each service class is service pricing which determines the revenue that the system receives when the service is rendered. The service provider would like to maximize the total revenue obtained by the system by means of optimal pricing for service classes and performing admission control functions subject to the bandwidth resources available in the system. The system achieves total revenue maximization first in a distributed manner by maximizing each cell's revenue for all service price combinations with the consideration of QoS constraints and then in a centralized manner by choosing the best price combination which generates the maximum total revenue for the system. That is, for possible service price combinations, each cell makes admission control decisions for new and handoff call requests in order to maximize the revenue received from servicing new and handoff calls in the cell. At the end, each cell generates a pricing-revenue table to be used by a central entity to determine best pricing that maximizes the total system revenue while satisfying QoS.

The optimal pricing and the total revenue obtained by the system are inherently related to the pricing algorithm employed by the service provider. While many pricing algorithms exist [10], the most prevalent with general public acceptance to date is the "charge-by-time" scheme by which a user is charged by the amount of time in service. We assume that such a "charge-by-time" pricing scheme is adopted by the service provider such that a call of service class i has a "charge-rate" of v^i per time unit. That is, if a call of service class i is admitted into a cell, and subsequently handed off to the next cell or terminated in the cell, a reward of v^i multiplied with the amount of time the service is rendered in the cell will be "earned" by the system. There is no distinction for handoff vs. new calls in pricing as long as the call is in the same service class. As suggested by empirical studies [1,12,15] pricing changes of each service class affect demands received for these services. We assume that QoS requirements are not affected by pricing change as indicated in [2]. Thus, the determination of optimal pricing v^i to each service class is part of revenue optimization. The performance model developed in the paper allows a service provider to calculate the revenue earned per unit time under an admission control algorithm by each individual cell such that the revenue obtained by the system is maximized while satisfying QoS constraints.

3. Methodology

We consider a "charge-by-time" pricing scheme with the goal to maximize the revenue obtainable by each cell while satisfying the QoS requirements. Many empirical studies have been done to determine the relationship between price and demand of telecommunication services. Our methodology to determine optimal pricing is applicable to any generic function of the form $\lambda = f(v)$ to relate demand changes with price changes. Without loss of generality, we consider the following pricing-demand function [1,12,15]:

$$\lambda^i = a^i (v^i)^{-\varepsilon^i}, \tag{1}$$

where λ^i and λ^i denote the arrival rate and pricing of service i while a^i and ε^i are constants correlating λ_i and v^i. The elasticity constant ε^i determines the effect of pricing changes on service demand. The demand of consumers toward a product depends on the affordability and necessity of the product. If the increase in demand is slower than the decrease in pricing, consumers are regarded as *inelastic* to price changes. On the other hand, if the increase in demand is faster than the decrease in pricing of the product, consumers are considered *elastic* toward the product. In Eq. (1), the elasticity constant ε^i indicates the elasticity of consumers to pricing changes of calls in service class i. A value greater than 1 predicts that lower pricing would generate higher arrival rates, while a value less than 1 predicts higher pricing would generate higher arrival rates. The elasticity ε^i value can be determined by analyzing statistical data collected for each service class. However, ε^i should be greater than 1 for most service classes as suggested by [1,12,15] reflecting the fact that lower pricing would stimulate higher demand or higher arrival rates for the service provided, and consequently generate higher revenue. The proportionality constant a^i differs from cell to cell and can be calculated from Eq. (1) once the reference arrival rates (λ^i), the current price (v^i) and the elasticity constant ε^i are known through statistical data.

For the case in which two service classes exist, the total revenue R_T generated by each cell is the sum of the revenue generated from each service class:

$$R_T = R_h^1 + R_n^1 + R_h^2 + R_n^2. \tag{2}$$

Here, R_h^1 represents revenue earned from servicing class 1 handoff calls, R_n^1 represents revenue earned from servicing class 1 new calls, and so on. Here, we note that pricing of a service class implicitly determines the arrival rate of that service class based on Eq. (1), which in turn affects the revenue obtainable by the system. If we lowered pricing of a service class, the arrival rate of that service class would increase. Since all service classes in a cell share bandwidth channels, if we lowered pricing of all service classes with the intent to increase revenue, QoS requirements would be violated because of system overload. Thus, the search for optimal prices to maximize the system revenue while satisfying QoS is a combinatorial problem.

Our approach is to calculate the maximum revenue obtainable as a result of applying a call admission control algorithm when given five parameter values for each service class i, namely, λ_n^i, $\mu_n^i\lambda_h^i$, μ_h^i and v^i, and use the revenue obtained as the objective function to guide the search process. Since pricing changes under the "charge-by-time" scheme are often incremental so there are not too many possibilities to search for the optimal combination. Hence, our approach is to exhaustively search all possible combinations of $[v^{i,min}, v^{i,max}]$ for all service classes and look for the best combination of service class prices that would maximize the system revenue while satisfying QoS. Specifically for each service class S^i, we obtain a 5-tuple "reference" parameter values $(\lambda_n^i, \mu_n^i\lambda_h^i, \mu_h^i, v^i)$ as the reference state. Then we determine the pricing range $[v^{i,min}, v^{i,max}]$ for the service class such that $v^{i,min}$ and $v^{i,max}$ are determined as minimum and maximum pricing acceptable by the customer base as apprised by the service provider. We then divide the pricing range $[v^{i,min}, v^{i,max}]$ into β^i parts, resulting in $\beta^i + 1$ potential prices for class i to be evaluated. We then take a price combination of potential prices for all classes and deduce the corresponding arrival rates of service classes based on Eq. (1) with respect to the reference state. The optimal pricing problem essentially corresponds to finding the best price combination that will maximize the system revenue with QoS guarantees.

Specifically in the first step, for each service class S^i where $i \in \{1, \ldots, n\}$, we determine $\beta^i + 1$ prices with equal increment of δ^i between $v^{i,min}$ and $v^{i,max}$ by using the following formula:

$$v^{i,j} = v^{i,min} + j \cdot \delta^i, \quad i \in \{0, \ldots, n\} \text{ and } j \in \{0, \ldots, \beta^i\}, \tag{3}$$

where

$$\delta^i = (v^{i,min} - v^{i,max})/\beta^i. \tag{4}$$

Basically, Eqs. (3) and (4) determine $\beta^i + 1$ potential prices, namely, $v^{i,min}$, $v^{i,j} + \delta^i$, $v^{i,min} + 2\delta^i, \ldots, v^{i,min} + (\beta^i - 1)\delta^i$ and $v^{i,max}$ for class i. We apply this step to all classes. Therefore, the total number of possible price combinations for all service classes from which to search for optimal pricing, denoted by η, is equal to

$$\eta = (\beta^1 + 1)(\beta^2 + 1) \ldots (\beta^n + 1). \tag{5}$$

In the second step, we predict the arrival rates of service classes for a given price combinations. Based on Eq. (1), the predicted new call arrival rate of the j^{th} price increment for S^i is given by

$$\lambda_n^{i,j} = a^i \cdot \left(v^{i,j}\right)^{-\varepsilon^i}. \tag{6}$$

Here, we note that ε^i remains the same while a^i differs from one cell to another. Since the user mobility is not affected by pricing changes, the ratio of new call arrival rates to handoff call arrival rates remains a constant. Thus, the predicted handoff call arrival rate of the j^{th} price increment for S^i is given by

$$\lambda_h^{i,j} = f^i \cdot \lambda_n^{i,j}, \tag{7}$$

where f^i denotes the ratio of the statistically determined reference handoff call arrival rate to the reference new call arrival rate for S^i, given by

$$f^i = \lambda_h^{i,current} / \lambda_n^{i,current}. \tag{8}$$

Note that f^i is a constant because handoff requests coming into each cell are modeled as the sum of the handoff requests coming into the reference cell from neighboring cells. Eq. (7) holds as long as the arrival rate increase or decrease in neighboring cells generates proportional increase or decrease on the handoff rate coming into the reference cell from these cells.

For all possible combinations, each of which generating a new 5-tuple $(\lambda_n^{i,j}, \mu_n^i\lambda_h^{i,j}, \mu_h^i, v^{i,j})$, we determine the revenue generated under a call admission control algorithm and store all the revenue values obtained in an n-dimensional revenue table where n denotes the number of service classes. This calculation procedure is done periodically by each cell. Tables in all the cells are collected and merged to determine global optimal pricing that would maximize the overall system revenue (sum of revenue from individual cells). In Section 4, we discuss how to utilize admission control algorithms integrated with pricing for revenue optimization to build such a table in individual cells based on this methodology. Later in Section 5, we discuss how these tables can be integrated to find global optimal pricing.

4. Admission control integrated with pricing for revenue optimization with QoS guarantees

Here, we give a brief overview of partitioning and threshold-based call admission control algorithms integrated with pricing as well as a hybrid partitioning-threshold algorithm for revenue optimization with QoS guarantees. We utilize these algorithms to calculate system revenue obtainable and determine optimal pricing based on the methodology discussed in Section 3. For ease of presentation, we assume that there are two service types, class 1 (high-priority) and class 2 (low-priority), distinguished primarily by their traffic type, i.e., real-time vs. non-real-time. Our methodology can be easily applied to the case in which more than two service classes exist. The traffic input parameters to our algorithm are λ_n^1, μ_n^1, λ_h^1 and μ_h^1 for class 1 and λ_n^2, μ_n^2, λ_h^2 and μ_h^2 for class 2. A call admission control algorithm adopted (among the three) will be executed by each individual cell to generate the system revenue obtainable while satisfying QoS constraints expressed in terms of B_h^1t, B_h^2t, B_n^1t, and B_n^2t, when given charge-by-time pricing of v^1 and v^2 for classes 1 and 2.

4.1. Partitioning admission control

A partitioning call admission control policy divides the total number of channels in a cell into several fixed partitions with each partition specifically reserved to serve a particular service class (real-time vs. non-real-time) and call type (new vs. handoff). In our system, the total number of channels, C, is divided into C_h^1, C_n^1, C_h^2, and C_n^2 channels for high-priority handoff calls, high-priority new calls, low-priority handoff calls, and low-priority new calls, respectively. Let $(n_h^1, n_n^1, n_h^2, n_n^2)$ be the numbers of calls corresponding to the four fixed partitions denoted by $C_h^1, C_n^1, C_h^2, C_n^2)$. Then $n_h^1k^1 = C_h^1$, $n_n^1k^1 = C_n^1$, $n_h^2k^2 = C_h^2$, and $n_n^2k^2 = C_n^2$ such that $C_h^1 + C_n^1 + C_h^2 + C_n^2 = C$. The optimization problem for the partitioning algorithm is to identify the best partition $(C_h^1, C_n^1, C_h^2, C_n^2)$ that would maximize the cell's revenue while satisfying the imposed QoS constraints defined by

$$B_h^1 < B_h^1t, B_n^1 < B_n^1t, B_h^2 < B_h^2t, B_n^2 < B_n^2t, \tag{9}$$

where B_h^1 and B_h^2 are the call dropping probabilities for handoff calls for various classes 1 and 2, respectively, and B_n^1 and B_n^2 are the blocking probability for new calls for classes 1 and 2, respectively.

A partitioning solution is "legitimate" if Condition 9 is satisfied. Since no sharing is allowed among partitions, the system behaves as if it is managing four concurrent subsystems, each of which behaves like an $M/M/n/n$ queue. The call dropping probabilities for handoff calls for various service classes (i.e., B_h^1 and B_h^2) and the blocking probability for new calls for various service classes (i.e., B_n^1 and B_n^2) can be determined easily by calculating the probability of the partition allocated to serve the specific calls being full. We can calculate the revenue generated per unit time by the partition reserved to serve only high-priority handoff calls by associating a reward of $i * v^1$ for state i in the $M/M/n_h^1/n_h^1$ queue. The same way applies to other partitions. Specifically, we can compute the revenue per unit time to the cell by $PR(C, \lambda_h^1, \lambda_n^1, \lambda_h^2, \lambda_n^2) = PR_h^1 + PR_n^1 + PR_h^2 + PR_n^2$, where the notation $PR(C, \lambda_h^1, \lambda_n^1, \lambda_h^2, \lambda_n^2)$ stands for the revenue earned by the partitioning algorithm as a function of $C, \lambda_h^1, \lambda_n^1, \lambda_h^2, \lambda_n^2$ (with other parameters not listed), while PR_h^1, PR_n^1, PR_h^2, and PR_n^2 stand for the revenues generated per unit time due to high-priority handoff calls, high-priority new calls, low-priority handoff calls, and low-priority new calls, respectively, as given by (only PR_h^1 is shown below since expressions for others are similar)

$$PR_h^1 = \sum_{i=1}^{n_h^1} i v^1 \frac{\frac{1}{i!}\left(\frac{\lambda_h^1}{\mu_h^1}\right)^i}{1 + \sum_{j=1}^{n_h^1} \frac{1}{j!}\left(\frac{\lambda_h^1}{\mu_h^1}\right)^j}. \tag{10}$$

The optimal partition $(C_h^1, C_n^1, C_h^2, C_n^2)$ is the one that maximizes $PR(C, \lambda_h^1, \lambda_n^1, \lambda_h^2, \lambda_n^2)$ and can be computed fairly easily by exhaustively searching through all the combinations that satisfy $C_h^1 + C_n^1 + C_h^2 + C_n^2 = C$. For a modern computer running Pentium 4, this takes only a few seconds for $C = 80$ channels. This is attributed to the existence of close-form solutions for computing B_h^1, B_n^1, B_h^2, B_n^2, PR_h^1, PR_n^1, PR_h^2, and PR_n^2 under partitioning admission control.

4.2. Threshold-based admission control

In the threshold-based admission control algorithm, we select a threshold C_T to separate class 1 from class 2 based on the service type, i.e., real-time vs. non-real time. The meaning of the threshold is that when the number of channels used in the cell exceeds C_T then new or handoff calls from service class 2 (low-priority) will not be admitted. Within each service class, we further create thresholds to differentiate handoff from new calls such that C_{hT}^1 is the threshold for class 1 high-priority handoff calls; C_{nT}^1 is the threshold for class 1 high-priority new calls; C_{hT}^2 is the threshold for class 2 low-priority handoff calls; and C_{nT}^2 is the threshold for class 2 low-priority new calls. Since we give handoff calls a higher priority than new calls, the following additional conditions must also be satisfied, $C_{nT}^1 \geqslant C_T$, $C_{hT}^1 \geqslant C_T$, $C_{nT}^2 \leqslant C_T$, and $C_{hT}^2 \leqslant C_T$. A threshold-based admission control integrated with pricing for revenue optimization with QoS guarantees thus aims to find the optimal set $(C_{hT}^1, C_{nT}^1, C_{hT}^2, C_{nT}^2)$ satisfying the above conditions that would yield the highest revenue with QoS guarantees.

The threshold-based admission control algorithm can be analyzed by using a SPN model [4] to compute B_h^i and B_n^i of class i. A "legitimate" solution from a threshold admission control algorithm must generate B_n^1, B_h^1, B_n^2, and B_h^2 to satisfy the QoS constraints specified by Condition 9. We compute the revenue generated per unit time from the threshold-based admission control algorithm to the cell by: $TR(C, \lambda_h^1, \lambda_n^1, \lambda_h^2, \lambda_n^2) = TR_h^1 + TR_n^1 + TR_h^2 + TR_n^2$. Here, TR_h^1, TR_n^1, TR_h^2, and TR_n^2 stand for the revenues generated per unit time due to high-priority handoff calls, high-priority new calls, low-priority handoff calls, and low-priority new calls, respectively, given by: $TR_h^i = (1 - B_h^i)\lambda_h^i v^i/\mu_h^i$, and $TR_n^i = (1 - B_n^i)\lambda_n^i v^i/\mu_n^i$.

The optimal threshold set $(C_{hT}^1, C_{nT}^1, C_{hT}^2, C_{nT}^2)$ is the one that maximizes $TR(C, \lambda_h^1, \lambda_n^1, \lambda_h^2, \lambda_n^2)$ and can be computed by searching through all the combinations that satisfy $C_{nT}^1 \geqslant C_T$, $C_{hT}^1 \geqslant C_T$, $C_{nT}^2 \leqslant C_T$, and $C_{hT}^2 \leqslant C_T$. For a modern computer running Pentium 4, this takes a few minutes for $C = 80$ channels. It takes a longer computation time compared with partitioning admission control because there is no close-form solution for B_h^1, B_n^1, B_h^2, B_n^2, TR_h^1, TR_n^1, TR_h^2, and TR_n^2 and it requires evaluating the SPN performance model developed [4] to generate the blocking probabilities and the revenue obtainable by the system with QoS guarantees.

4.3. Hybrid partitioning-threshold admission control

The hybrid partitioning-threshold admission control algorithm takes advantage of both partitioning and threshold-based. The hybrid algorithm divides the channels into fixed partitions the same way as the partitioning algorithm does. In addition, to take advantage of multiplexing, a "shared" partition is reserved to allow calls of all service classes/types to compete for its usage in accordance with the threshold algorithm. The shared partition is available for use by a service class/type only if the partition reserved for that service class/type is used-up. Let $n_{hs}^1, n_{ns}^1, n_{hs}^2, n_{ns}^2$ be the numbers of high-priority handoff calls, high-priority new calls, low-priority handoff calls, and low-priority new calls, respectively, in the shared partition. Let C_s be the number of channels allocated to the shared partition under the hybrid algorithm. Then, the number of calls of various service classes and types admitted into the shared partition are limited by C_s channels allocated to the shared partition, that is, $n^1_{hs}k^1 + n^1_{ns}k^1 + n^2_{hs}k^2 + n^2_{ns}k^2 \leqslant C_s$ subject to the constraint that $C^1_h + C^1_n + C^2_h + C^2_n + C_s = C$. The QoS constraints are specified by Condition 9.

Note that the hybrid algorithm encompasses the partitioning algorithm as a special case in which $C_s = 0$ and also the threshold-based algorithm as another special case in which C^1_h, C^1_n, C^2_h, and C^2_n are all zero. The performance model for the hybrid algorithm is composed of two sub-models: one for the partitioning algorithm with the four fixed partitions C^1_h, C^1_n, C^2_h, and C^2_n and one for the threshold-based algorithm for which $C = C_s$. Since the fixed partitions are modeled as M/M/n/n queues, the arrival rates into the shared partition from high-priority handoff calls (λ^1_{hs}), high-priority new calls (λ^1_{ns}), low-priority handoff calls (λ^2_{hs}), and low-priority new calls (λ^2_{ns}) are simply the spill-over rates from their respective $M/M/n/n$ queues, e.g.,

$$\lambda^1_{hs} = \lambda^1_h \frac{\frac{1}{n^1_h!}\left(\frac{\lambda^1_h}{\mu^1_h}\right)^{n^1_h}}{1 + \sum_{j=1}^{n^1_h} \frac{1}{j!}\left(\frac{\lambda^1_h}{\mu^1_h}\right)^j}. \tag{11}$$

Here, only λ^1_{hs} is shown since expressions for λ^1_{ns}, λ^2_{hs}, and λ^2_{ns} are similar. From the perspective of the shared partition, the arrival rates are thus λ^1_{hs}, λ^1_{ns}, λ^2_{hs}, and λ^2_{ns} and the total number of channels available is C_s with all other parameters remained the same. Hence, we compute the revenue generated per unit time from the hybrid admission control algorithm to the cell by the sum of revenue earned from the fixed partitions plus that from the shared partition, i.e., $HR(C, \lambda^1_h, \lambda^1_n, \lambda^2_h, \lambda^2_n) = PR(C - C_s, \lambda^1_h, \lambda^1_n, \lambda^2_h, \lambda^2_n) + TR(C_s, \lambda^1_{hs}, \lambda^1_{ns}, \lambda^2_{hs}, \lambda^2_{ns})$. The optimization problem for the hybrid algorithm is to identify the best partition $(C^1_h, C^1_n, C^2_h, C^2_n, C_s)$ and the best threshold $(C^1_{s\,hT}, C^1_{s\,nT}, C^2_{s\,hT}, C^2_{s\,nT})$ within C_s that would maximize the cell's revenue subject to the imposed QoS constraints defined by Condition 9. It could be found by searching through all the combinations that satisfy $C^1_h + mC^1_n + C^2_h + C^2_n + C_s = C$ and $C^1_{s\,nT} \geqslant C_{sT}$, $C^1_{s\,hT} \geqslant C_{sT}$, $C^2_{s\,nT} \leqslant C_{sT}$, and $C^2_{s\,hT} \leqslant C_{sT}$ where $C_{sT}$is a threshold used within the shared partition C_s, $C_{sT} \leqslant C_s$, to separate class 1 from class 2. For a modern computer running Pentium 4, this takes minutes to search for the best solution for $C = 80$ channels. Again there is no close-form solution for $B^1_h, B^1_n, B^2_h, B^2_n$, and HR and it requires evaluating the SPN performance model developed [4] to generate the blocking probabilities and the revenue obtainable by the system with QoS guarantees.

5. Numerical analysis

We present numerical data for $\eta = 6 \times 8 = 48$ possible future price combinations by applying the revenue formulas derived for partitioning, threshold-based and hybrid admission control algorithms. These possible future price combinations are relative to current pricing (of the reference system) such that the price increment/decrement is considered acceptable to the service provider. We compare performance characteristics of these admission control algorithms with QoS guarantees with physical interpretations given. The analysis considers two classes. Class 1 (real-time) demands more resources with higher QoS constraints than class 2 (non-real-time). Thus, class 1 has more stringent call blocking probabilities than class 2, as well as higher pricing. We consider the call arrival process for each class follows the Poisson distribution which has been frequently used in the literature to model call arrivals. Thus, the inter-arrival time of service calls is exponentially distributed. We note that because we use the SPN model for performance evaluation we can also accept general time distributions to relax this assumption.

The input parameters are C, λ^1_h, μ^1_h, λ^1_n, μ^1_n, λ^2_h, μ^2_h, λ^2_n, μ^2_n, v^1, v^2, a^1, a^2, ε^1, ε^2, k^1, k^2, $B^1_h t$, $B^2_h t$, $B^1_n t$, and $B^2_n t$. We set $C = 80$ channels, $k^1 = 4$ and $k^2 = 1$ for a typical cell in mobile wireless networks to service real-time and non-real-time traffic such that there are 80 channels in the cell with a class 1 call (real-time) consuming 4 channels and a class 2 call (non-real-time) consuming 1 channel. We assume that the statistical data collected for the reference cell would provide $\lambda^1_h = 5$, $\mu^1_h = 1$, $\lambda^1_n = 2$, $\mu^1_n = 1$, $\lambda^2_h = 4.4$, $\mu^2_h = 1$, $\lambda^2_n = 4.4$, $\mu^2_n = 1$ per minute, and the current pricing is $v^1 = 80$ cents/min, $v^2 = 12$ cents/min. Similarly, we set elasticity constants to values greater than 1, $\varepsilon^1 = 1.3$ and $\varepsilon^2 = 1.7$ for class 1 and class 2 calls, respectively. We apply Eq. (6) to calculate proportionality constants $a^1 = 600$ and $a^2 = 300$ for class 1 and class 2 calls, respectively, for our reference cell. We vary service prices in the range $[50, 100]$ for v^1 and $[6, 20]$ for v^2 and the resulting call arrival rates are calculated by using Eqs. (6) and (7).

Figs. 1–3 show the maximum revenue obtained by partitioning, threshold-based and hybrid admission control algorithms, respectively. Class 1 and class 2 prices are shown on the Y and X coordinates, respectively, in the unit of cents/min. The maximum revenue obtainable by a legitimate solution is shown on the Z coordinate also in the unit of cents/min.

Fig. 1 indicates that the revenue obtainable increases as the anticipated arrival rate increases as a result of lowering the prices, as long as the QoS constraints can still be satisfied. Nevertheless, as λ^1_h and λ^1_n increase to 2.4 and 6 for $v^1 = 70$, the partitioning admission control algorithm fails to yield a legitimate solution because the workload is too heavy to satisfy the imposed QoS constraints. Likewise as λ^2_h and λ^2_n increase to 8.7 when $v^2 = 8$, the algorithm fails to yield a legitimate solution. The maximum revenue at 664 is established at $v^1 = 80$ and $v^2 = 10$ (the highest point in Fig. 1) as these prices result in the highest arrival rate that can be handled with QoS guarantees. The best partition reserved to handle the traffic generated for $(v^1 = 80, v^2 = 10)$ is $(C^1_h = 10, C^1_n = 5, C^2_h = 11, C^2_n = 9)$ while the best partition reserved to handle the traffic generated in the current system for $(v^1 = 80, v^2 = 12)$ is $(C^1_h = 10, C^1_n = 5, C^2_h = 10, C^2_n = 10)$. As the arrival rate of class 2 becomes higher due to the price cut of v^2 from 12 to 10, partitioning admission control in this case reduces C^2_n and increases C^2_h to satisfy the higher QoS requirement of class 2 handoff calls.

Fig. 2 shows that the highest revenue at 722 is achieved under threshold admission control when $v^1 = 80$ and $v^2 = 6$ (the highest point in Fig. 2). To satisfy QoS requirements of class 1 calls, the system applies thresholds $C^1_{nT} = 80, C^1_{hT} = 80, C^2_{nT} = 76, C^2_{hT} = 76$ to

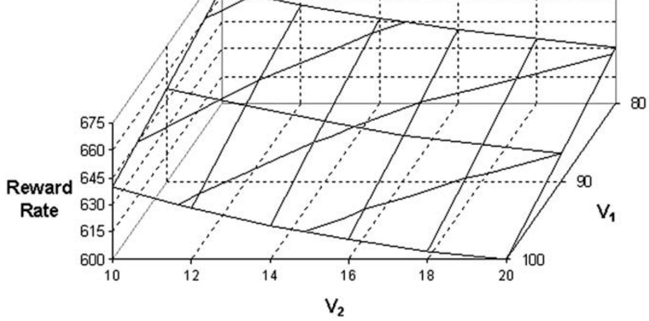

Fig. 1. Maximum revenue obtained by partitioning admission control algorithm.

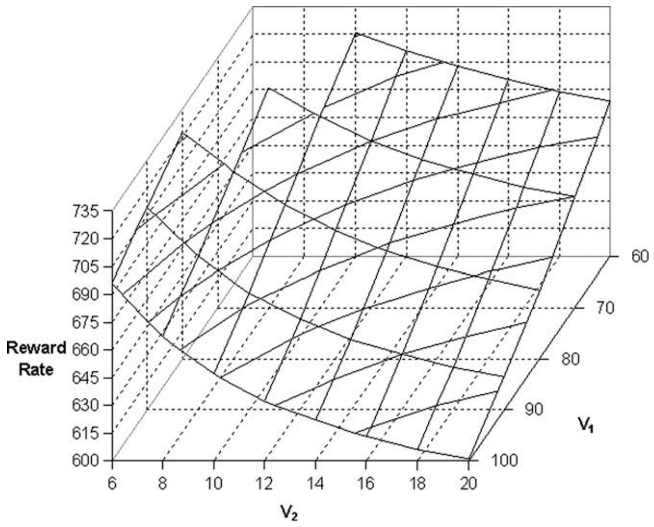

Fig. 2. Maximum revenue rate obtained by threshold admission control algorithm.

handle higher class 2 traffic generated for ($v^1 = 80, v^2 = 6$), as opposed to all thresholds being set to 80 for the current system with ($v^1 = 80, v^2 = 12$). By sharing resources among service classes and controlling the effect of higher class 2 arrival rates by applying lower threshold values, threshold-based admission control performs better than partitioning admission control.

Fig. 3 illustrates the maximum revenues obtainable with hybrid admission control with QoS guarantees as a function of v^1 and v^2. The highest revenue at 736 is achieved when $v^1 = 60$ and $v^2 = 8$ (the highest point in Fig. 3). Recall that hybrid admission control reserves ($C_h^1 = 7, C_n^1 = 3, C_h^2 = 1, C_n^2 = 1, C_s = 38$) to handle the traffic generated for the reference system with ($v^1 = 80, v^2 = 12$). To handle higher class 1 and class 2 arrival rates due to optimal pricing at ($v_1 = 60$ and $v_2 = 8$), it reserves ($C_h^1 = 6, C_n^1 = 1, C_h^2 = 1, C_n^2 = 0, C_s = 51$) and applies a lower threshold to class 2 calls in the common partition. In response to a higher class 1 and class 2 arrival rates, hybrid admission control tends to increase the size of C_s partition.

Comparing Fig. 3 with Figs. 1 and 2, it is clearly seen that the hybrid admission control algorithm outperforms both partitioning and threshold based algorithms. The multiplexing power of the shared partition is clearly demonstrated by the fact that hybrid always significantly outperforms partitioning in terms of revenue obtainable over a wide range of class 1 and class 2 service call arrival values, while being able to sustain a higher workload and provide QoS guarantees. We observe that the performance of threshold-based admission control is comparable to hybrid admission control until both class 1 and class 2 arrivals become very high ($\lambda_h^1 = 7.3, \lambda_n^1 = 2.9, \lambda_h^2 = 8.7, \lambda_n^2 = 8.7$ anticipated when $v_1 = 60$ and $v_2 = 8$). At these high arrival rates, threshold-based admission control fails to yield a legitimate solution compared with hybrid admission control. We attribute the superiority of hybrid admission control over partitioning and threshold-based admission control to the ability to optimally reserve dedicated resources for high-priority classes through fixed partitioning to reduce interference from low-priority classes, and to optimally allocate resources to the shared partition in accordance with threshold-based admission control to exploit the multiplexing power for all classes. We conclude that the channel allocation made by the hybrid admission control algorithm represents the best possible way to handle higher arrival rates and to allow a wider ranger of pricing for revenue optimization. Also hybrid admission control generates maximum revenue obtainable while satisfying QoS requirements for the example cell.

To apply the results obtained in the paper, each cell would independently collect statistical data periodically to estimate a set of reference arrival and departure rates of new/handoff calls of various service classes based on statistical analysis [4]. Each cell then determines the proportionality constant a^i for each service class by applying Eq. (1) based on current pricing, the arrival rate, and the elasticity constant ε^i of each service class. Later each cell determines new/handoff call arrival rates for a range of "future" potential pricing for each service class also based on Eq. (1). Finally, for each candidate price combination, each cell calculates optimal $(C_h^1, C_n^1, C_h^2, C_n^2, C_s)$ and optimal $(C_s^{1_{hT}}, C_s^{1_{nT}}, C_s^{2_{hT}}, C_s^{2_{nT}})$ within C_s if the hybrid admission control algorithm is used for revenue optimization with QoS guarantees assuming two classes exist. The optimal settings for all future price combinations are then summarized in a revenue table and reported to a central entity which collects and analyzes revenue tables from all the cells in the system. To guarantee QoS, the particular future price combination that satisfies QoS constraints in all of the cells while maximizing the aggregate revenue would be chosen as the winner for optimal pricing.

The system provider can have each cell generate such a revenue table only periodically, e.g., every 3 months, as deemed economically feasible by the service provider for changing pricing to service classes. When the end of the current period is approaching, a new

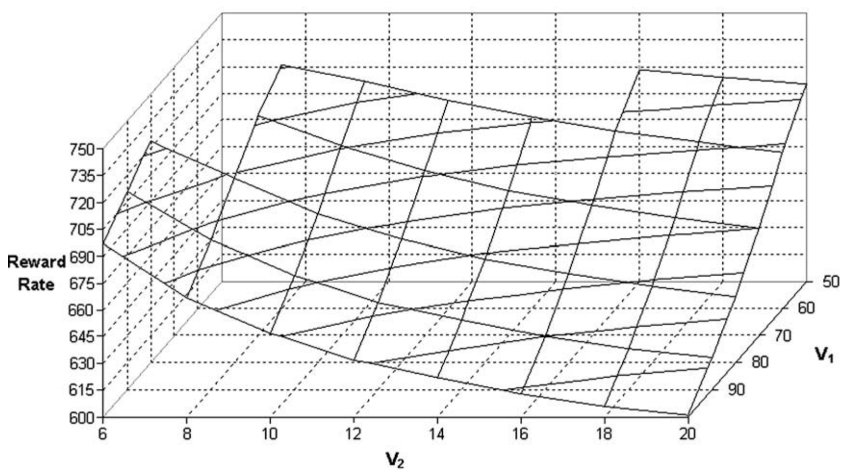

Fig. 3. Maximum revenue obtained by hybrid admission control algorithm.

set of reference arrival and departure rates of new/handoff calls of various service classes that have been collected over the last period will be used as input to build the revenue table by each cell. Thus, the revenue table can be built by each cell in the background periodically. With C=80 channels in a cell, the computational time taken for a cell to regenerate the revenue table is less than 1 day when using a modern Pentium 4 machine. The system-wide optimal pricing per service class then can be determined by the service provider using the methodology introduced here. Such optimal pricing determined will then remain static across cells for all service classes till the next period.

The above approach is distributed in nature and will only require the central entity to collect revenue tables reported by individual cells, and just pick the price combination that satisfies QoS constraints in all of the cells while maximizing the aggregate revenue. Another approach is based on centralized control by having the central entity guide the search such that cells will be instructed to evaluate a particular price combination only when needed until the best price combination is found. This approach has the advantage that each cell does not regenerate the entire revenue table.

Finally the service provider may elect to have different pricing in different days of the week (e.g., weekday vs. weekend rate) or even in different times of the day (e.g., day-time vs. night-time rate). In this case, the same methodology developed in the paper applies, except that it is being applied separately to each time segment. For example, if there is a distinction of weekend vs. weekday rate, then two separate sets of reference arrival and departure rates of new/handoff calls would be collected for weekend and weekday time segments, and used as input to generate optimal pricing separately for these two time segments.

6. Summary

In this paper, we proposed and analyzed a methodology to determine optimal pricing for revenue optimization with QoS guarantees in wireless mobile networks, utilizing admission control algorithms integrated with pricing. Our methodology is based on the idea that the maximum revenue generated by a cell while satisfying QoS depends on both the admission control algorithm chosen by the system and pricing applied to each service class. To determine optimal pricing, we first applied an empirical function that relates pricing with demand so as to predict the change to the arrival rate of a service class when its price changes. Then we tested a range of future pricing for each of the multiple service classes, each combination of which generates a new set of demand arrival rates as input to feed into a call admission control algorithm to calculate the revenue generated with QoS guarantees as the objective function. We discovered that a hybrid scheme combining the benefits of both partitioning and threshold-based performs the best in terms of revenue maximization with QoS guarantees and optimal pricing that maximizes the revenue earned.

This work is only a beginning of the design concept of utilizing admission control algorithms cognizing revenue optimization with QoS guarantees to determine optimal pricing in wireless networks. Possible future research directions extending from this work include (a) designing and analyzing a search algorithm to guide the search for the best price combination so as to avoid having each cell regenerate the entire revenue table periodically; (b) considering other pricing schemes (flat rate or charge by connection) and investigating optimal resource allocation settings under which hybrid admission control can yield the highest revenue with QoS guarantees; (c) considering other revenue collection model, e.g., revenue is collected on call termination or revenue is lost when a call is terminated prematurely.

References

[1] M. Aldebert, M. Ivaldi, C. Roucolle, Telecommunications demand and pricing structure: an economic analysis, in: Proc. of the 7th Conf. Telecommunications Systems: Modeling and Analysis, Nashville, TN, 1999, pp. 255–267.

[2] A. Bouch, M.A. Sasse, Why value is everything: a user centric approach to internet quality of service and pricing, IWQoS, Karlsruhe, Germany, June 2001.

[3] I.R. Chen, C.M. Chen, Threshold-based admission control policies for multimedia servers, The Computer Journal 39 (9) (1996) 757–766.

[4] I.R. Chen, O. Yilmaz, I.L. Yen, Admission control algorithms for revenue optimization with QoS guarantees in mobile wireless networks, Wireless Personal Communications 38 (3) (2006) 357–376.

[5] Y. Fang, Thinning algorithms for call admission control in wireless networks, IEEE Transactions on Computers 52 (5) (2003) 685–687.

[6] R. Guerin, Queuing-blocking systems with two arrival streams and guarded channels, IEEE Transactions on Communication 36 (1988) 153–163.

[7] D. Hong, S.S. Rappaport, Traffic model and performance analysis for cellular mobile radio telephone systems with prioritized and non-prioritized handoff procedures, IEEE Transactions on Vehicular Technology VT35 (3) (1986).

[8] D. Hong, S.S. Rappaport, Priority oriented channel access for cellular systems serving vehicular and portable radio telephones, Communications, Speech and Vision, IEE Proceedings I 131 (5) (1989) 339–346.

[9] J. Hou, J. Yang, S. Papavassiliou, Integration of pricing with call admission control to meet QoS requirements in cellular networks, IEEE Transactions on Parallel and Distributed Systems 13 (9) (2002) 898–910.

[10] N.J. Keon, G. Anandalingam, Optimal pricing for multiple services in telecommunications networks offering quality-of-service guarantees, IEEE/ACM Transactions on Networking 11 (1) (2003) 66–80.

[11] M. Kyriakakos, N. Frangiadakis, L. Merakos, S. Hadjiefthymiades, Enhanced path prediction for network resource management in wireless LANs, IEEE Wireless Communications 10 (6) (2003) 62–69.

[12] S. Lanning, D. Mitra, Q. Wang, M. Wright, Optimal planning for optical transport networks, Philosophical Transactions Royal Society London. Series A 358 (1773) (2000) 2183–2196.

[13] B. Li, C. Lin, S.T. Chanson, Analysis of a hybrid cutoff priority algorithm for multiple classes of traffic in multimedia wireless networks, Wireless Networks 4 (1998) 279–290.

[14] Y.B. Lin, I. Chlamtac, Wireless and Mobile Network Architecture, Wiley, New York, 2001.

[15] P. Rappaport, J. Alleman, L.D. Taylor, Household demand for wireless telephony: an empirical analysis, in: Proc. of the 31st Annual Telecommunications Policy Research Conf., Arlington, VA, September 2003.

[16] J. Wang, Q. Zeng, D.P. Agrawal, Performance analysis of a preemptive and priority reservation handoff algorithm for integrated service-based wireless mobile networks, IEEE Transactions on Mobile Computing 2 (1) (2003) 65–75.

[17] J. Ye, J. Hou, S. Papavassiliou, A comprehensive resource management for next generation wireless networks, IEEE Transactions on Mobile Computing 1 (4) (2002) 249–263.