An Approach to Connection Admission Control in Single-hop Multi-service Wireless Networks with QoS Requirements

Tara Javidi and Demosthenis Teneketzis
Department of Electrical Engineering and Computer Science
University of Michigan
Ann Arbor, MI 48109
taraj@eecs.umich.edu teneket@eecs.umich.edu

Abstract

We formulate a resource allocation problem in single-hop multi-service networks with quality of service requirements. We present a decomposition of the problem into two analytically tractable subproblems. We illustrate the approach for the case where the QoS requirement is expressed in terms of outage probability. We establish a sufficient condition for the optimality of the greedy policy in the above resource allocation problem.

I. Introduction-Motivation

The scarcity of available resources, such as limited bandwidth and low capacity, as well as the interference among users result in serious challenges in the design for wireless networks. An important network layer design problem is the efficient allocation of those limited resources. An efficient allocation must optimize a performance objective while satisfying the quality of service (QoS) required by each type of service and connection (expressed in terms of signal to noise plus interference ratio (SNIR), and outage probability, latency, etc.).

In this paper we present a systematic approach to connection admission control (CAC) in single-hop multi-service wireless networks with QoS requirements. The approach consists of the decomposition of the resource allocation problem into two sub problems:(i) the specification of an admission region $\mathcal{A}$ which guarantees the QoS requirements for each connected user, independently of the admission policy; and (ii) the determination of a connection admission policy that is optimal within the class $\Pi_{\mathcal{A}}$ of policies restricted to the admission region $\mathcal{A}$. such a decomposition results in tractable problems and creates a conceptual framework for understanding the interaction among the different layers of the wire less communication networks.

The remainder of the paper consists of four parts. In Part 1, we (i) formulate the CAC in single hop multi-service wireless networks with QoS requirements; (ii) discuss the nature of this problem and the need for alternative tractable methods to solve it; and (iii) propose the aforementioned decomposition. In Part 2, we (i) present an approach to defining probability of outage as a system-wide QoS measure for cellular systems; and (ii) construct an admission region $\mathcal{A}$ where, independently of admission strategies, the requirements on probability of outage are satisfied. In Part 3, we address the CAC problem under a constraint described by the admission region $\mathcal{A}$. In Part 4, we present conclusion and reflections.

Part 1. Outage-based Admission region

In a wireless system the desirable resource allocation is achieved through two separate mechanisms of power-rate assignment (PRA) and connection admission control (CAC). In systems where rate of transmission of any user and the power control mechanism is fixed and known, CAC is the only mechanism to guarantee a certain level of service while maximizing total revenue over horizon T. Mathematically the question can be formulated as $\max_\pi E\left\{\sum_{t=0}^{T}\sum_{i=1}^{L} c_i x_i^\pi(t)\right\}$ such that $q_k^i(\pi) = \theta_k(\mathbf{x}^\pi(1),\dots,\mathbf{x}^\pi(T)) \in Q_i$; where c_i is the rate of revenue generated by a user of type i, $x_i^\pi(t)$ is the number of connections of type i present at the system at time t and is a function of CAC strategy π, $\mathbf{x}^\pi(t)$ is the vector of number of users $(x_1^\pi(t), x_2^\pi(t),\dots,x_L^\pi(t))$, $q_k^i(\pi)$ is a vector of QoS measures for user k of type

i under CAC strategy π which is a function of the sequence of $\{x_i^{\pi}(t)\}_{t=1}^{T}$. θ_k is such function and its form depends on the physical layer and the power assignment rule.

Notice that quality of service for a connection is a dynamic variable whose statistics depend on the chosen admission strategy. Hence, a key feature of this optimization problem is that there is a two-way coupling between the constraints resulting from the QoS requirements and the admission policy. Such a two-way coupling results in a computationally challenging and analytically intractable optimization problem.

In this paper we propose the two step decomposition described in the introduction. Such a decomposition results in a one-way coupling between the constraints present in the resource allocation problem and the determination of an optimal allocation policy. Though, in general, our approach results in a suboptimal solution for the original problem, it reduces the complexity of the problem to a great extent. Furthermore, it creates a conceptual framework for understanding the interaction among different layers of wireless communication systems, such as physical layer concerns, QoS requirements, and network layer resource allocation. In other words, the admission region $\mathcal{A} = \{\mathbf{x} : q_k^i(\pi) \in Q_i for \forall \pi \in \Pi_{\mathcal{A}}\}$ conceptualize the physical channel and QoS requirements; and the optimization problem is reduced to $\max_{\pi \in \Pi_{\mathcal{A}}} E\left\{\sum_{t=0}^{T}\sum_{i=1}^{L} c_i x_i^{\pi}(t)\right\}$

Part 2. Outage-based Admission region

Outage probability is an important performance measure in cellular networks. In a cellular scenario low SNIRcan increase bit error rate, but more importantly if this ratio remains low for a long enough duration, it can cause an outage in an ongoing service (due to loss of synchronization, etc). This will result in disconnection of an admitted call. In most common scenarios, this is considered a more severe form of low performance than blocking (which occurs when a new call is denied admission to the cell, hence the network). As a result, outage probability is considered a main perfomance measure for traditional cellular networks.

We describe an outage by two parameters: (i) the SNIR threshold γ; and (ii) a minimum duration τ. An outage occurs when the SNIR remains below the threshold γ for a period longer than or equal to τ. In most of the currently available literature (e.g. see [9], [12]), an outage is assumed to occur when the SNIR falls below a threshold γ^*. We believe that this is not sufficient to capture the essence of an outage, since it ignores statistical correlation or burstiness in the incoming traffic stream. It is intuitively expected that traffic streams with high level of burstiness are more probable to cause an outage than non-bursty or iid streams with the same level of instantaneous instantaneous interference. Similarly, the memory present in shadowing channels directly affects how long the impairment will last, hence it affects the occurance of an outage. In other words, the drop in the SNIR below γ does not result in an outage instantaneously; an outage results in when the SNIR is low for an extended period of time, i.e. a time period that exceeds a minimum duration τ. With this definition, the occurance of outage events strictly depend on the second order statistics of the interference and/or shadowing. A characterization of outage both in terms of the threshold γ and the time duration τ has appeared only in [10] and [20]. One key feature of [10] and [20] is that the effect of other users on the outage probability is not taken into account. That is, the effect of the (random) number of active users and the statistical variation of their channels on the probability of outage is ignored. Attention in both [10] and [20] is restricted on one user and on the effect of its physical channel on the outage probability. In general, the performance of a wireless system critically depends on two factors: (i) the condition of the physical channel; and (ii) the interference created by other users. Indeed, we show that by incorporating the effect of multiple access interference into our approach, we are able to relate the outage probability to the number and type of users present in the system and, therefore, to determine an admission region associated with the maximum acceptable outage probability for each type of users.

The salient features of our approach are the following: $(F1)$ We model the statistical variation of the physical channel by a Markov Chain (as in [20]). $(F2)$ We consider several types of users in terms of their statistical activity, and QoS requirements. $(F3)$ We fix the total number of users admitted by the system, and we assume that the status of each user switches between "active" and "inactive" according to a Markov rule (independent of $(F1)$). The status of a particular user is not necessarily independent of that of another user.

As a result of the aforementioned features, we can construct a model which allows us to define, for any multiple access scheme, the SNIR ratio and hence, determine for any parameters γ and τ the probability of outage as a function of the fixed number of users present in the system. This in turn allows us to analytically determine the capacity of the system (described in terms of an admission region) associated with maximum acceptable probability of outage. Therefore, we achieve two main goals in this part of the paper: (i) the development of an approximate statistical model for outage and calculation of the outage probability; and (ii) the analytic determination of an admission region based the on the desired performance of the system with regard to outage probability.

This part is organized as follows: In Section 2.I, we construct a stochastic model,analytically calculate the probability of outage, and provide a procedure to construct an outage-based admission region. In Section 2.II we present examples illustrating the modeling and results in Section 2.I.

I. Outage-based Admission Region for Multiuser Systems with Markov Channels

A. Philosophy of Our Approach

We address the issue of outage within the context of QoS requirements. A user in the system encounters an outage event when its received SNIR at the base station falls below a threshold for an extended period of time. Hence, an outage is experienced by each user individually. Therefore, the key conceptual issue is how to analytically describe an outage event as a system-wide QoS criterion. We address this issue by introducing a fictitious observer/user and by defining an outage incurring during this user's service time. To guarantee that the outage-based QoS requirements are satisfied for every type of user that may be admitted by the system we proceed as follows: We consider a separate fictitious observer/user for each type of traffic. Such a user is always active and is identical to the actual users of the same type in terms of the statistics of the physical channel, SNIR threshold, and minimum outage duration. Each fictitious observer/user does not create any interference in the system, hence has no effect on the performance of the system. The outage probability for such a user is a conservative bound on the outage probability of each user of the same type. The system-wide QoS requirement in terms of outage probability is met if and only if the probability or the frequecy of outage for each of the aforementioned fictitious users is below a prespecified value (that depends on the type of user) which reflects the QoS requirement. In this section, we construct the outage-based admission region following the above philosophy.

B. Outage Formulation for a Given Observer/User in the Presence of a Fixed Number of Users

We fix the number of admitted users, and then develop an approach to defining and computing outage probability for a fictitious observer/user u_0, whose channel statistics, SNIR threshold γ, and minimum outage duration τ are given.

In a wireless setting the received SNIR of an observer/user u_0 depends on two decoupled factors: 1) the effect of physical channel in the absence of other users; this captures events like additive noise, fading, and/or shadowing (in the presence or absence of power-control mechanisms). 2) the effect of the presence, power, and channel statistics of the other active users admitted in the system. Therefore, to determine the probability of outage, we need: (i) to model the channel degradation; (ii) to model the interference of other admitted users; and (iii) to construct a "Super Markov Chain" combining (i) and (ii) in order to describe the received SNIR of u_0. For a detailed description of a model that completely accounts for the effect of these phenomena see [6]. In this paper we consider only the worst case scenario for u_0, where the channel of all the other users are in their best realization. Furthermore, we assume that users of similar type with similar channel realization are assigned similar transmission power. In this situation the received SNIR of u_0 depends simply on the effect of the physical channel and the presence of other active users.

It is very common to model the effect of the channel on SNIR in the absence of other users as a Markov chain. The validity of such model has been extensively studied and confirmed in the literature (see [19]). The most commonly used example of this kind is the Gilbert Channel. In general, such a MC is defined by its state-space $\mathcal{H} = \{h_1, h_2, \ldots, h_I\}$, and its transition matrix $\mathbf{A} = [a_{kl}] := [\text{Prob}\{X(t+1) = h_l | X(t) = h_k\}]$. Note that in the case of an ideal power control mechanism, the state-space $\mathcal{H}$ is reduced to a singleton $\{h\}$, hence $\mathbf{A} = 1$; in case of power control with quantized error $\pm\delta$, we have $\mathcal{H} = \{h - \delta, h, h + \delta\}$. We assume that channel states of individual users are mutually independent.

To model the interference of other admitted users, we assume that there are L types of users in terms of QoS requirements, transmission Power, and the activity factor [14], and there are $(M_1, M_2, \cdots, M_L)$ users admitted to the system (not including u_0). At any time slot, each admitted user can be active ("on") or inactive ("off"). Since only active users interfere with the received signal of u_0, we need to find an appropriate model to describe the evolution of the users' "on" periods. In this paper, we assume that active and inactive periods for a user of type l evolve according to a b_l-order Markov chain. Consequently the number of type l active users can be modeled by a Markov chain whose state is denoted by an integer $r_l \in \{0, 1, \ldots, M_l\}$. In general we assume that the activity of all users can be correlated. Based on the above we can express the state of the number of active users by the the random vector $(r_1, r_2, \ldots, r_L)$. By construction, this array evolves according to a known Markov rule. Let $\mathbf{T}$ be the transition matrix for this Markov chain, i.e. $\mathbf{T} = [t_{ij}] := [\text{Prob}\{(r_1, r_2, \ldots, r_L) = j \mid (r_1, r_2, \ldots, r_L) = i\}]$. Note that T is a square matrix of dimension $\prod_{l=1}^{L} (M_l + 1)$.

To describe the received SNIR of u_0, we construct a "super Markov chain" (SMC) which represents the variation of the physical channel for u_0 and the number of the admitted users. The states of this SMC are vectors of form $(h^{u_0}, r_1, r_2, \cdots, r_L)$. where $h^{u_0} \in \mathcal{H}_{u_0}$ is the state of the channel between user u_0 and the base-station, and r_l, $l = 1, 2, \ldots, L$, as mentioned before, denotes the number of type l active user. The state-space of this SMC is $\mathcal{S} = \mathcal{H}_{u_0} \times \prod_{l=1}^{L} \{0, 1, 2, \ldots, M_l\}$. Since by assumption, the state of the physical channel for a user is independent of the number of the other users and their channel state, the transition probability for this SMC can be easily obtained by

$$\mathbf{P} = \mathbf{T} \otimes \mathbf{A}_0 \tag{1}$$

where $\mathbf{A}_0$ is the transition matrix of the Markov physical channel between observer/user u_0 and the base-station, and $A \otimes B$ denotes the Kronecker product of the matrices A and B. If we denote by I_{u_0}, the size of set $\mathcal{H}_{u_0}$, then the size of matrix $\mathbf{P}$ is $I_{u_0} \times \prod_{l=1}^{L} (M_l + 1)$.

To define an outage event mathematically, we must specify the received SNIR of observer/user u_0 at each state $\mathbf{s} \in \mathcal{S}$. This SNIR is a function $f_{u_0} : \mathcal{S} \longmapsto \mathbb{R}_+$. The exact form of $f_{u_0}(\cdot)$ depends on the dynamics of multiple access interference, and possibly the power control mechanism. For instance, for a CDMA system where users of the same class have a common transmitted power and there is no power control, the form of function f_{u_0} is:

$$f_{u_0}(\mathbf{s}) = \frac{h_0^u P_0 G_0}{\eta + \sum_{l=1}^{L} h_{best}^l P_l r_l} \tag{2}$$

where η is the noise power (that includes the expected total interference from the adjacent cells), r_l is the the total number of active users of type l, G_0 is the spreading gain for user u_0, h^{u_0} is the channel gain between u_0 and the base station, and P_l is the common transmitted power for all type-l users. This form can extend to CDMA systems with power control where each class of users has a common targeted power, and where $h^{l,k}$ represents the error of the power control mechanism.

After specifying the SNIR of user u_0 at each state, we define the set of "bad states" $\mathcal{B}$ as

$$\mathcal{B} := \{\mathbf{s} \in \mathcal{S} | f_{u_0}(\mathbf{s}) < \gamma\}. \tag{3}$$

Based on the above classification of states we can now formally define the following:

Definition 1: An outage is an event where the state of the SMC enters $\mathcal{B}$ and stays in $\mathcal{B}$ for at least τ units of time.

Definition 2: The probability of an outage is defined to be the probability that a randomly selected time slot belongs to an outage event and is denoted by P_{outage}.

C. *Outage Analysis for a Given Observer/User in the Presence of a Fixed Number of Users*

The probability and frequency of an outage event in the constructed SMC can be studied in the framework of [20].

Consider the constructed SMC and the associated transition matrix $\mathbf{P}$ with it. We follow [20] to establish the necessary equations and relations that describe the probability of outage. Note that the SMC is mathematically equivalent to the physical Markov Channel studied in [20], even though the SMC, in general, has a much larger state space, and it has a very specific structure due to its construction. Hence, after introducing the appropriate notation and definitions we can use results provided from [20] for the analysis of the probability of outage.

C.1 Definitions

- Let the row-vector $\boldsymbol{\pi}$ denote the stationary distribution of SMC.
- Define $\boldsymbol{\pi_G}$ as

$$\boldsymbol{\pi_G}(\mathbf{s}) = \begin{cases} \boldsymbol{\pi}(\mathbf{s}) & \text{if } \mathbf{s} \in \mathcal{G} \\ 0 & \text{otherwise} \end{cases}, \tag{4}$$

- Define $\boldsymbol{P_\mathcal{B}}$ as the matrix with entries

$$\boldsymbol{P_\mathcal{B}}(i,j) = \begin{cases} \mathbf{P}(i,j) & \text{if } j \in \mathcal{B} \\ 0 & \text{if } j \in \mathcal{G} \end{cases} \tag{5}$$

C.2 Results

We establish an analytical expression for probability of outage. For that matter we need the following result from [20].

Fact. 1: The probability of outage is given as

$$P_{outage} = \tau \boldsymbol{\pi_G} \boldsymbol{P_\mathcal{B}}^{\tau} \mathbf{1} + \boldsymbol{\pi_G} \boldsymbol{P_\mathcal{B}}^{\tau+1} (I - \boldsymbol{P_\mathcal{B}})^{-1} \mathbf{1} \tag{6}$$

Based on Fact 1 we establish an alternative analytical expression for the probability of outage. The new expression is easier to compute as it involves neither inversion of a matrix nor calculation of vector $\boldsymbol{\pi_G}$. For the proof of Proposition 1 see [6].

Proposition 1: $P_{outage} = \boldsymbol{\pi} \boldsymbol{P_\mathcal{B}}^{\tau} \left(\tau I - (\tau - 1) \boldsymbol{P_\mathcal{B}}\right) \mathbf{1}$

D. *Construction of an Outage-Based Admission Region*

We now discuss how to use the results obtained in Section 2.I-C to construct an admission region when the probability of outage is the QoS requirement under consideration. An admission region is the set of all combinations of admitted users such that if connection admissions are restricted to a subset of its interior, the probability of an outage encountered by a fictitious observer/user of type l is less than a prespecified threshold P^l_{max} for all $l = 1, 2, \ldots, L$.

The formulation of probability of outage presented in Section 2.I-B and the analysis of Section 2.I-C provide an expression for the probability of outage of a fictitious user of type l ($l = 1, 2, \ldots, L$) as a function $g_l : \mathbb{Z}^L_+ \longmapsto \mathbb{R}_+$ of the vector of admitted users $\underline{M} = (M_1, M_2, \ldots, M_L)$. Therefore, for a fixed type l fictitious user (i.e. $\mathcal{H}_{u_0} = \mathcal{H}_l$, $\mathbf{A}_0 = \mathbf{A}_l$, $\tau_0 = \tau_l$, $\gamma_0 = \gamma_l$, and $f_{u_0}(\cdot) = f_l(\cdot)$) the region where QoS (expressed by the probability of outage) is guaranteed for that type of user is

$$\mathcal{R}_l := g_l^{-1}([0, P^l_{max}]) = \{\underline{M} : g_l(\underline{M}) \leq P^l_{max}\}. \tag{7}$$

Consequently, the region where the QoS is guaranteed for all the users is

$$\mathcal{R} := \bigcap_{l=1}^{L} \mathcal{R}_l = \bigcap_{l=1}^{L} \{\underline{M} : g_l(\underline{M}) \leq P^l_{max}\}. \tag{8}$$

Since it is not desirable for any admission strategy to terminate an unfinished service, we define the admission region $\mathcal{A}$ as the largest coordinate convex subset of $\mathcal{R}$, i.e.

$$\mathcal{A} := \sup_{\mathcal{C} \subseteq \mathcal{R}} \{\mathcal{C} : \text{ if } \underline{M} \in \mathcal{C}, \underline{M}' \leq \underline{M} \text{ then } \underline{M}' \in \mathcal{C}\}. \tag{9}$$

Recall that our analysis is valid for a fixed number of admitted users. In a wireless system the number of users (active and inactive) present in the system varies with time. We establish the validity of our analysis for wireless systems through the following theorem (For the proof see [6]).

Theorem 1: The admission region $\mathcal{A}$ is a conservative bound on the number of admitted users for which the QoS expressed by the probability of outage under any admission strategy is met.

II. Special Cases, Examples, Discussion

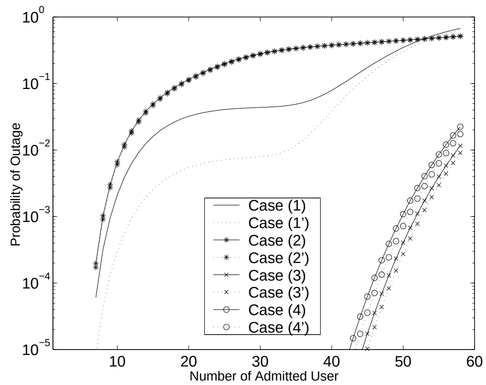
In this section we present examples illustrating our approach. In all the examples we consider the outage problem in a CDMA pre-third-generation wireless systems. In such systems, traffic mainly consists of voice, or data streams that are compressed and then treated as voice [16]. This kind of traffic, when the number of users is fixed, can be appropriately modeled by an Engseth birth-death chain (see [1]).

In the remainder of the section we formally introduce the Engseth traffic model. In Sections 2.II-A we compute the probability of outage, and the resulting admission regions under a variety of power control scenarios and system parameters.

We consider L types of traffic. Let the component M_k of the vector $\underline{M} = (M_1, \ldots, M_L)$ represent the fixed number of users of type k admitted to the system. Let $\underline{N}(t) = (N_1(t), \ldots, N_L(t))$ denote the vector of the number of active users of each type at time t. Then the transition probability for the Engseth model is given as follows:

$$P\{\underline{N}(t+1) = \boldsymbol{n} | \underline{N}(t) = \boldsymbol{m}, \underline{M}\} = \begin{cases} (M_k - m_k)\lambda_k & \text{if } n = m + e_k \\ m_k \mu_k & \text{if } n = m - e_k \\ 1 - \sum_{k \in K_{on}} m_k \mu_k & \\ - \sum_{k \in K_{off}} (M_k - m_k)\lambda_k & \text{if } n = m \\ 0 & \text{otherwise} \end{cases}$$

where e_k is a column vector whose elements are all zero except for the k^{th} element which is 1, λ_k is the activation rate of each inactive user of type k, μ_k is the probability of that an active



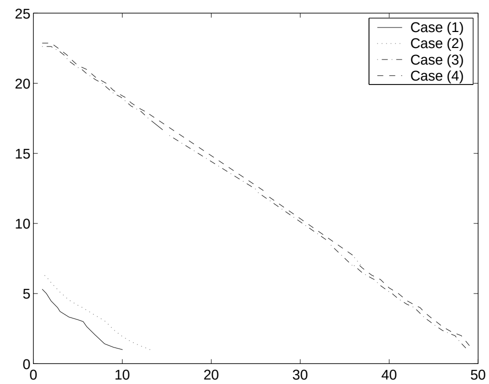


Fig. 1. Left) P_{outage} vs. M, for cases in 2.II-A. Right) Admission Regions for cases in 2.II-A.

user becomes inactive, $K_{off} = \{k : m_k < M_k\}$, and $K_{on} = \{k : 0 < m_k\}$. Note that $\frac{\lambda_k}{\lambda_k+\mu_k}$ is the activity factor of each stream. For voice users, this is around 0.4. For data users, it varies with the application, and it depends on the burstiness and information bandwidth of the stream, as well as the compression method employed.

A. Numerical Examples

We first construct the SMC associated with this model. Let M denote the number of admitted users in the system. We construct $\mathbf{T}_M = [\text{Prob}\{r = i/r = j, M\}]$.

For the cases that follow we use different channel models $(\mathcal{H}, \mathbf{A}_0)$. We construct the transition probability $\mathbf{P}$ associated with each $(\mathcal{H}, \mathbf{A}_0)$ pair, using (1). After identifying the transition matrix, determining the SNIR at each state, and finally labelling "bad" states according to (3), we calculate the probability of an outage for type l connections as a function $g_l(M)$.

We first study the homogeneous traffic scenarios. For all the scenarios under study, we assume that $\frac{\lambda}{\lambda+\mu} = 0.4$, $T = .002\tau_f$ (T is the time slot duration and τ_f is the fading cycle), $\gamma = 3.1$dB, and the spreading gain is $G_0 = 64$. Fig. 1. Left) shows the result of such a calculation for CDMA systems when: 1) the channel follows a Gilbert model with average burst lengths of 4, with steady-state probability of the bad-channel-state equal to 0.1 and $\tau = 7T$ (the value recommended by ITU-T [13]); 1') channel is similar to one of (1), and $\tau = 15T$; 2) the channel is an appropriate approximation to a Rayleigh fading channel with the maximum Doppler frequency of 100 Hz as given by [19] and $\tau = 7T$; 2') the channel is similar to (2) and $\tau = 15T$; 3) an ideal power control mechanism is implemented and $\tau = 7T$; 3') the channel is similar to (3) and $\tau = 15T$; and 4) power control is applied with error of 5% and $\tau = 7T$. 4') the channel is similar to (4) and $\tau = 15T$;

Now we study the outage problem for the same CDMA system when the traffic consists of two classes of users with different activity factors, spreading gains, and outage parameters; these parameters are $\frac{\lambda_1}{\lambda_1+\mu_1} = 0.4$, $\frac{\lambda_2}{\lambda_2+\mu_2} = 0.6$, $G_1 = 64$, $G_2 = 32$, $P_2 = 2P_1$ $\tau_1 = \tau_2 = \tau$, $\gamma_1 = 3.1$dB, and $\gamma_2 = 3.3$dB. We set the maximum acceptable probability of outage to be equal to 10^{-3}. Under this specification, Fig. 1. Right) shows the admission region, when: 1) the channel is described by a Gilbert model similar to the one in Section 2.II-A and $\tau = 7T$; 2) the channel is described by a Gilbert model similar to the one in Section 2.II-A and $\tau = 15T$; and 3) there is an ideal power control mechanism and $\tau = 7T$; and 4) there is an ideal power control mechanism and $\tau = 15T$.

B. Discussion

Fig. 1 illustrates that for all cases discussed in Section 2.II-A $g_1(M)$ is an increasing function of M. Similar plots are provided in [6]. These plots, like Fig. 1, show that $g_l(M_1, M_2)$, $l = 1, 2$, is increasing in M_1 and M_2. This implies that the region $\mathcal{R}$ defined by (8) is coordinate convex, hence $\mathcal{A} = \mathcal{R}$ for the examples studied. Based on these result we propose the following conjecture:

Conjecture 1: In any cellular system, $g_l(\underline{M})$ is increasing in each coordinate M_l for all $l = 1, 2, \ldots, L$; Hence, $\mathcal{A} = \mathcal{R}$.

Part 3. Connection Admission Control

As discussed in the introduction, We propose to formulate the Connection Admission Control (CAC) problem in a single-hop multi-service network with QoS requirements as a constrained stochastic dynamic optimization problem, where the constraint describes the admission region. One standard approach to describing the admission region for such a problem is to define a total capacity for the network and associate an effective bandwidth to each class of users. This approach approximates the boundary of the admission region with a linear function of the number of each type of users (see [3], [17], and [18]). In cases where all the QoS requirements are summarized by an effective bandwidth-based admission region, the CAC problem is equivalent to a classical knapsack problem. (see [15] chapters 2-4, and the references therein for details on the classical stochastic knapsack problem).

In general, all QoS requirements considered simultaneously are summarized by an admission region the boundary of which need not be a line. In such a situation a reasonable assumption on the nature of the admission regions and their boundaries is coordinate convexity, which implies that no forced termination of service is required in order to meet QoS requirements. Based on this observation, in this paper we propose the formulation and investigation of a "generalized knapsack" whose scheduling is equivalent to the CAC problem with a coordinate convex admission region.

To motivate the analysis of the CAC problem presented in Section 3.I we first briefly discuss and critique the results available on the classical knapsack problem. Several variants of the classical knapsack problem have been carefully studied in literature (for example see [2], [4],[5], [7], [8], [11], and [15]. In [11] reward that is fixed and known, and is obtained at the instance of admission. This feature makes the problem considered in [11] distinctly different from the problem we consider in this paper. In [8] it is assumed that no job admitted to the knapsack leaves the system, i.e. the problem changes to a packing problem. Such a problem is also distinctively different from ours. The model and the formulation of the knapsack problem considered in [2], [5], [7], and [15, Ch. 4] are similar to our problem. A Markov Decision Process (MDP) approach is used in this class of references for the analysis of the classical knapsack problem. It has been shown that solving the appropriate MDP, through standard numeric programming methods, can be analytically intractable and computationally complex. Furthermore, it is known that the optimal solution to such MDP (which is the optimal admission policy for the classical knapsack), in general, lacks any specific structure or well-definable property (See [7]).

The complicated nature of the optimal connection admission control policies create a practical difficulty for their implementation as viable CAC policies in high speed networks. Consequently, in this paper we consider the "greedy policy" which has a very simple implementation. The greedy policy admits any request for connection if the resulting number and configuration of admitted users belongs to set $\mathcal{A}$. We determine conditions on the rates of revenue generated by different classes of connections sufficient to guarantee the optimality of the greedy policy. The problem we address can be thought of as follows: How should each type of service provided by the network be charged so that it should be optimal to admit every request for connection provided that there are sufficient resources?

The remainder of this part of paper is organized as follows: In Section 3.I we formulate the CAC problem with two classes of connections as a generalized knapsack problem, called Problem **(P)**. We provide a sufficient condition on the optimality of greedy policy for Problem

(P) (for the proof, see [6]). Section 3.II includes a brief discussion of a further extension of the CAC problem.

I. The Generalized Stochastic Knapsack Problem with Two Classes of Connections

The generalized stochastic knapsack problem with two classes of users can be formulated as follows:

*Problem **(P)***

Consider a finite coordinate-convex set $\mathcal{S} \subset \mathbb{R}_+^2$ which contains the origin. A two dimensional generalized knapsack associated with set $\mathcal{S}$, consists of a system of identical servers in parallel that can serve two classes of service within its support region $\mathcal{S}$. That is, the knapsack may serve x_1 number of class-1 connections, and x_2 of class-2 connections, only if $(x_1, x_2) \in \mathcal{S}$. Each connection of class k, $k = 1, 2$, is characterized by its arrival rate, λ_k, and the rate of its service time, μ_k. We assume that: arrival and service statistics of each connection are independent of each other and independent of the arrival and service statistics of other connections; the service time for a connection of type k is a memoryless random variable with mean $\frac{1}{\mu_k}$; each unit of time there is at most one new connection arrival to the system. Each arriving connection can be admitted to the knapsack if the resulting number of connections is in $\mathcal{S}$. If a request for connection is rejected, the connection is lost. An admitted connection remains in the knapsack until its service is completed. Without any loss of generality and for clarity, we assume that arrivals and departures within the time slot from time t to $t+1$ occur in the open interval $(t, t+1)$; furthermore, departures occur at the end of a time slot whereas arrivals occur at the beginning of a time slot. Thus, if we define t^+ and t^- as

$$\begin{aligned} t^+ &= \inf\{s \in (t, t+1) : s \text{ is the time after the arrival time of new} \\ &\qquad \text{connection requests and admission decisions in time slot } t\} \\ (t+1)^- &= \sup\{s \in (t, t+1) : s \text{ is the time before the completion time} \\ &\qquad \text{of any connection whose service ends in time slot } t\}. \end{aligned}$$

we have $t^+ < (t+1)^-$. Each admitted connection of type $k = 1, 2$ generates a revenue of rate c_k while being served in the knapsack. The goal is to find an optimal admission strategy that maximizes the total expected revenue over horizon T, where T may be infinite.

Remark: As a result of our formulation, a packet of type k may be admitted in the system at t^+, complete service at $(t+1)^-$, and result in a revenue c_k.

The main result of this section is summarized by the following theorem:

Theorem 2: If

$$\frac{\lambda_2}{\mu_2} \le \frac{c_1}{c_2} \le \frac{\mu_1}{\lambda_1}, \tag{10}$$

then the policy that follows the greedy rule at all times is optimal for Problem **(P)**.

For the analysis of Problem **(P)** and proof of Theorem 2, see [6].

II. Extensions and Generalizations

A. Generalized Knapsack Problem L ($L > 2$) Classes of Connections

It is possible to establish, by arguments similar to those used in the proof of Theorem 2, the following result for the generalized knapsack problem with L types of connections.

Theorem 3: Consider Problem **(P)** with L types of users. If

$$c_l \ge \sum_{k \neq l} c_k \frac{\lambda_k}{\mu_k}, \qquad \forall l \in \{1, 2, \ldots, L\} \tag{11}$$

then the policy that follows the greedy rule at all times is optimal.

We note that as L increases the sufficient conditions, described by (11), for optimality of the greedy admission policy become increasingly weak.

Part 4. Conclusion

In this paper we presented an approach to the connection admission control for a single-hop multi-service wireless network with QoS requirements. In general, a connection admission control strategy creates a complicated two-way coupling between the physical layer, i.e. QoS, and the network layer, i.e. the optimal resource allocation. Our approach proposes a decomposition of the problem in two subproblems: admission region construction and generalized knapsack scheduling. The result of such decomposition is reducing the interaction of the two layer into a one-way coupling between the physical layer (QoS) and the network layer (CAC). To demonstrate the methodology, we, then, constructed an outage-based admission region. Simultaneous consideration of QoS requirements such as outage probability, average bit error rate, delay, etc., can be incorporated into the admission control problem by taking the intersection of the corresponding admission regions resulting from the above QoS requirements. Such an intersection defines the admission region for a generalized knapsack problem. We investigated a generalized knapsack problem and established conditions sufficient to guarantee the optimality of the greedy admission policy.

ACKNOWLEDGMENT

This research was supported in part by ARO Grant DAAH04-96-1-0377 and NSF Grant ECS-9979347.

REFERENCES

[1] S. Asmussen. *Applied Probability and Queues.* John Wiley & Sons, 1987.

[2] C. Barnhart, J. Wieselthier, and A. Ephremides. Admission-control policies for multihop wireless networks. *Wireless Networks*, 1:373–387, 1995.

[3] J. S. Evans and D Everitt. Effective bandwidth-based admission control for multiservice CDMA cellular networks. *IEEE Transactions on Vehicular Technology*, 48(1):36–46, January 1999.

[4] G. J Foschini, B. Gopinath, and J. F. Hayes. Optimum allocation of servers to two types of competing costumers. *IEEE Transactions on Communications*, 29(7):1051–1055, July 1981.

[5] A. Gavious and Z. Rosberg. A restricted complete sharing policy for a stochastic knapsack problem in B-ISDN. *IEEE Transactions on Communications*, 41(7):2375–2379, JULY 1994.

[6] T. Javidi and D. Teneketzis. Outage, QoS, and admission region in a single cell. Control Group Report CGR-01-07, University of Michigan, EECS Department, RM. 4230, EECS BLDG. Ann Arbor, MI 48109-2122 USA, March 2001.

[7] S. Jordan and P. P. Varaiya. Control of multiple service, multiple resource communication networks. *IEEE Transactions on Communications*, 42(11):2979–2988, November 1994.

[8] T. E. Lee and G. T. Oh. The assymptotic value-to-capacity ratio for the multi-class stochastic knapsack problem. *European Journal of Operational Research*, 103:584–594, 1997.

[9] J. Lin, W. Kao, Y. T. Su, and T. Lee. Outage and coverage consideration for micro-cellular mobile radio systems in a shadowed-Rician/shadowed-Nakagami environment. *IEEE Transactions on Vehicular Technology*, 48(1):66–75, January 1999.

[10] N. B. Mandayam, P. Chen, and J. M. Holtzman. Minimum duration outage for CDMA cellular systems: A level crossing analysis. *Wireless Personal Communication*, (7):135–146, 1998.

[11] S. Martello and P. Toth. *Knapsack Problems.* J. Wiley & Sons, 1990.

[12] S. Oh and K. M. Wasserman. Dynamic spreading gain control in multiservice CDMA networks. *IEEE Journal on Selected Areas in Communication*, 17(5):918–927, May 1999.

[13] R. O. Onvural. *Assynchronous transfer mode networks: Performance issues.* Artech House, 1994.

[14] T. S. Rappaport. *Wireless Communications: Principle & Practice.* Prentice Hall, 1996.

[15] K. W. Ross. *Multiservice Loss Models for Broadband Telecommunication Networks.* Springer, 1995.

[16] J. Sullivan and A. Mendelson. Personal communication services: Bringing new quality and clarity to the enterprise. InfoTech: PCS Reports 1, Phillips, Fall 1997.

[17] D. Tse, , and S. Hanly. Effective bandwidths in wireless networks with multiuser receivers. In *Proceedings of the 17th Annual IEEE Conference on Computer Communications (INFOCOM)*, volume 1, page 3542, 1998.

[18] D. Tse and S. V. Hanly. Linear multiuser receivers: Effective interference, effective bandwidth and user capacity. *IEEE Transactions on Information Theory*, 45(2):641–657, March 1999.

[19] H. S. Wang and N. Moayeri. Finite-state Markov chain: A useful model for radio communication channels. *IEEE Transactions on Vehicular Technology*, 44(1):163–171, February 1995.

[20] M. Zorzi. Outage and error events in bursty channels. *IEEE Transactions on Communications*, 46(3):349–356, March 1998.