

# Coordinated 4G Connection Access Control and Resource Allocation

Chao Yang and Scott Jordan  
University of California, Irvine  
Email: Chao.Yang@uci.edu, sjordan@uci.edu

**Abstract**—We investigate how connection access control and resource allocation in cellular systems can be coordinated to jointly achieve maximum total utility. We propose a decomposition in which resource allocation maximizes long-term average utility for each system state and connection access control maximizes long-term average utility over all system states. We discuss the resulting interface and gives examples of algorithms that satisfy this decomposition and interface. Performance is illustrated by numerical results.

## I. INTRODUCTION

Heretofore, connection access control (CAC) and resource allocation (RA) have been designed to accomplish different goals in cellular networks. Resource allocation algorithms typically attempt to maximize the total utility of all active users, e.g. the total number of voice users or the total throughput of data users. In contrast, connection access control algorithms typically admit a new connection if and only if it is believed that capacity is available to ensure acceptable performance. Thus while RA focuses on utility, CAC typically ignores utility and merely focuses on capacity.

Here, we investigate how CAC and RA can be coordinated to both focus on utility. There is ample reason to believe that coordination of CAC and RA could be beneficial. The volume of data traffic on 4G networks has surpassed that of voice, and video is quickly becoming the dominant traffic class by volume, and it is expected that voice, data and video will all be important revenue generators. Whereas the first three generations of cellular networks segregated capacity for voice and data applications, future networks will transmit all applications over the Internet Protocol (IP) and thereby share capacity amongst all application classes. Multiplexing of traffic classes with different Quality-of-Service (QoS) requirements presents a challenge to CAC and RA. When capacity was segregated, CAC and RA for voice could both focus on maximizing the number of voice users, and CAC and RA for data could both focus on maximizing throughput. In contrast, when capacity is shared, it no longer is meaningful for CAC to admit a new connection if and only if it is believed that capacity is available to ensure acceptable performance, since such an admit decision for a voice call may result in an unacceptably high marginal decrease in utility for active data and video connections. Instead, we suggest that CAC should also focus on maximizing utility.

There is a great deal of research literature on uncoordinated resource allocation and connection access control for cellular

networks. Resource allocation algorithms usually attempt to either maximize total throughput or total utility under QoS requirements, e.g. minimum SINR or minimum rate, see e.g. [1][2][3][4][5][6]. Connection access control has typically depended upon the application type. For voice, CAC usually admits a new call if and only if there are available resources, see e.g. [7]. In contrast, data applications are often treated as not having any QoS requirement, and thus not requiring any CAC. CAC for video applications has been less addressed and remains an open problem.

However, there is little research literature on coordinated resource allocation and connection access control for cellular networks. A number of papers propose a weak type of coordination in which CAC admits voice users subject to available capacity and RA attempts to maximize the total utility of data users subject to performance constraints on voice users, see e.g.[8]. Other papers propose coordination by treating the goal as one of maximizing sum throughput of voice and data subject to QoS requirements, see e.g. [9]. We are not aware of any literature that proposes CAC to maximize total utility in cellular networks.

Here, we propose coordinated CAC and RA on the basis of user utility. First, we suggest a joint optimization of long-term average user utility over both CAC and RA policies. Then, we propose a decomposition into separate CAC and RA problems. Whereas traditional RA in the literature typically attempts to maximize total utility over a relatively short time period, we propose that RA should attempt to maximize the *long-term average utility for each system state*, where the state is defined as number of active applications of each type. Whereas traditional CAC in the literature typically admits new connections if and only if there are available resources, we propose that CAC should attempt to maximize the long-term average utility *over all system states*. We discuss the resulting interface between RA and CAC. We also give examples of RA and CAC algorithms that satisfy this decomposition and interface.

In section II, we define a user's channel, rate, and utility. In section III, we formulate a joint CAC and RA problem and propose a decomposition into separate CAC and RA problems. In section IV, we provide an implementation example to explain how to evaluate the average utility of each state and how to use dynamic programming to design an connection access control policy. Finally, in section V the performance of our framework is illustrated by numerical simulation results.

## II. SYSTEM MODEL

We consider a single cell downlink Orthogonal Frequency-Division Multiplexing (OFDM) system with  $N$  subcarriers. The bandwidth  $B$  of each subcarrier is assumed to be less than the coherence bandwidth of the channel so that the channel response can be considered flat. The rate of user  $k$  on subcarrier  $n$  at time  $t$  is:

$$r_{k,n,t}(p_{k,n,t}) = B \log_2 \left( 1 + p_{k,n,t} \frac{|H_{k,n,t}|^2}{\sigma^2 + I} \right)$$

where  $p_{k,n,t}$  is the power allocated,  $|H_{k,n,t}|^2$  is the composite channel gain,  $\sigma^2$  is the noise power, and  $I$  is the interference power. The channel gain  $|H_{k,n,t}|^2 = \alpha_{k,n,t}^2 \gamma_{k,t} PL_{k,t}$  is composed of fast fading  $\alpha_{k,n,t}^2$  which changes significantly in sequential time periods, slow fading and shadowing  $\gamma_{k,t}$  which changes little in sequential time periods but may change significantly during a few seconds, and pathloss  $PL_{k,t}$  which depends on user position and changes significantly during tens of seconds. Fast fading on different subcarriers is assumed to be independent to each other. The total rate of user  $k$  at time  $t$  is the sum of the user's rate over all subcarriers:

$$R_{k,t} = \sum_{n=1}^N r_{k,n,t}$$

Assume there are  $L$  types of applications. Each application type is associated with a utility function  $U_l$  which measures user satisfaction with the average rate achieved within a time window  $W_l$ . Denote the application of user  $k$  by  $b_k = l$ , and user  $k$ 's average rate at time  $t$  within the most recent window  $W_{b_k}$  by  $S_{k,t} = \int_{\tau=t-W_{b_k}}^t R_{k,\tau} / W_{b_k} d\tau$ . User utility is thus a function of average rate  $U_{b_k}(S_{k,t})$ . We wish to consider elastic applications (e.g. data) modeled by concave utility, semi-elastic applications (e.g. video) modeled by sigmoid utility, and inelastic application (e.g. voice) modeled by step utility. The QoS requirements of inelastic and semi-elastic applications are modeled by a minimum rate requirement  $S'_l$  on application type  $l$ . Arrivals of applications of type  $l$  are assumed form independent Poisson processes with intensity  $\beta_l$ , and the duration of connections of type  $l$  are assumed to be independent and Exponentially distributed with mean  $1/\gamma_l$ .

## III. JOINT FRAMEWORK

In this section we first formulate a joint CAC and RA problem. Then, we propose a decomposition into separate CAC and RA problems, and discuss the resulting interface between CAC and RA.

### A. Joint CAC and RA Optimization Problem

We begin by consideration of the policy spaces for CAC and RA. For CAC, we assume that departures are never blocked, so that the policy space can be written as:

$$G = \{\mathbf{g} = (g_1, g_2, \dots, g_L) : g_l \in \{0, 1\}, l = 1, \dots, L\}$$

where  $g_l = 0$  (resp.  $g_l = 1$ ) denotes that an arrival of type  $l$  should be blocked (resp. admitted). Denote the time of the

$j$ th event (arrival or departure) by  $t^{(j)}$ , and the set of active users at time  $t$  by  $D_t = \{k | \text{user } k \text{ is active at time } t\}$ . For each arrival event  $j$ , the CAC policy is given by  $\mathbf{g}^{(j)} = \{g_1^{(j)}, g_2^{(j)}, \dots, g_L^{(j)}\}$ . The state of the network is represented by a vector  $\mathbf{x} = \{x_1, x_2, \dots, x_L\}$  where  $x_l$  is the number of active type  $l$  applications; the state just after event  $j$  is denoted by  $\mathbf{x}^{(j)} = \{x_1^{(j)}, x_2^{(j)}, \dots, x_L^{(j)}\}$ . The CAC policy, denoted by  $Q^{CAC} = \mathbf{g}^{(j)}(\mathbf{x}^{(j-1)})$ , thus decides on admission of a connection on the basis of the state immediately before the connection's arrival. The RA policy, denoted by  $Q^{RA} = \{p_{k,n,t} | |H_{k,n,\tau}|^2, \forall k, n, \tau \leq t \text{ and } \mathbf{x}^{(j)}\}$ , assigns powers to each user and subcarrier at each time  $t$  on the basis of each user's historical channels and on the current state.

We now turn to the desired optimization metric. Resource allocation algorithms typically attempt to maximize the total utility of all active users during a short time period and connection access control algorithms typically admit a new connection if and only if it is believed that capacity is available to ensure acceptable performance. Such uncoordinated RA and CAC has several disadvantages. First, if CAC does not consider utility, then admitting a new arrival if capacity allows may be sub-optimal, since the expected utility from the new arrival may well be less than the decrease in utility of current users resulting from their decrease in average future rate. Thus, we suggest that CAC should also focus on maximizing utility. Second, by focusing on utility, when residual capacity is low and a relatively low-paying application type arrives, CAC can judge whether it is optimal to block the arrival with the hope that higher-paying application may arrive soon. Finally, if RA focuses on short-term performance, then long-term utility is not necessarily maximized.

We thus propose that both RA and CAC should focus on long-term user utility. There are several options for this metric. Whereas traditional RA in the literature typically attempts to maximize total utility over a relatively short time period, we propose that RA should attempt to maximize the *long-term* average utility for each system state. Define the duration of state  $\mathbf{x}^{(j)}$  as  $\Delta t^{(j)} = t^{(j+1)} - t^{(j)}$ . Denote the average utility per unit time of state  $\mathbf{x}^{(j)}$  by:

$$U_{avg}(\mathbf{x}^{(j)}, \Delta t^{(j)}) = \frac{1}{\Delta t^{(j)}} \int_{t^{(j)}}^{t^{(j+1)}} \sum_{k \in D_t} U_{b_k}(S_{k,t}) dt$$

Note that this metric thus averages utility over different time periods in which the state is the same, as opposed to maximizing short-term throughput.

Whereas traditional CAC in the literature typically admits new connections if and only if there are available resources, we propose that CAC should attempt to maximize the long-term average utility *over all system states*. The joint connection access control and resource allocation problem is thus:

$$\max_{Q^{CAC}, Q^{RA}} \lim_{J \uparrow \infty} \frac{1}{t^{(J+1)}} \sum_{j=0}^J U_{avg}(\mathbf{x}^{(j)}, \Delta t^{(j)}) \Delta t^{(j)} \quad (1)$$

where  $J$  denotes the total number of events.

The same optimization metric is thus used by both CAC and RA. This formulation, however, requires the joint determination of the optimal CAC and RA policies. An exhaustive search amongst all possible CAC and RA policies within the joint policy space is almost certainly infeasible. In the next two subsections, we thus attempt to decompose the problem into separate CAC and RA problems. The challenge is to formulate decomposed problems that retain the desired coordination and that require a simple interface between the two.

### B. Decomposition and Interface

We start with resource allocation. Traditional resource allocation focuses on one period of time and optimizes system performance during this period. We argue that resource allocation should attempt to maximize the long-term average utility for each system state, i.e. there should be a RA policy for each state  $\mathbf{x}$ . In an infinite length of time period, state  $\mathbf{x}$  occurs during times

$$T(\mathbf{x}) = \bigcup_{\{j \leq J: \mathbf{x}^{(j)} = \mathbf{x}\}} [t^{(j)}, t^{(j+1)})$$

The average utility of state  $\mathbf{x}$  over an infinite time period is thus

$$U_{avg}(\mathbf{x}) = \lim_{J \uparrow \infty} \frac{1}{|T(\mathbf{x})|} \int_{T(\mathbf{x})} \sum_{k \in D_t} U_{b_k}(S_{k,t}) dt$$

where  $|\cdot|$  denotes cardinality.

For each state  $\mathbf{x}$ , resource allocation should maximize long-term average utility and the resource allocation problem can be written as:

$$\begin{aligned} \bar{U}_{avg}(\mathbf{x}) &= \max_{Q^{RA}} U_{avg}(\mathbf{x}) \\ \text{s.t.} \quad &\sum_{k \in D_t} \sum_{n=1}^N p_{k,n,t} \leq P, \quad \forall t \in T(\mathbf{x}) \\ &p_{k,n,t} \geq 0, \quad \forall k \in D_t, n, t \in T(\mathbf{x}) \\ &S_{k,t} > S'_{b_k}, \quad \forall k \in D_t, t \in T(\mathbf{x}) \end{aligned} \quad (2)$$

We define the feasible region  $\Lambda$  as the set of states for which there exists a resource allocation that satisfies the constraints in (2):

$$\Lambda = \{\mathbf{x} : \exists Q^{RA} \text{ which can solve problem (2)}\}$$

We turn next to connection access control. While RA attempts to maximize the long-term average utility for each state, CAC should attempt to maximize the long-term average utility over all states. The major question is: does CAC need detailed knowledge of the RA algorithm or can it treat RA as a black box? The literature doesn't provide a clear answer to this question. To strengthen the modularity of the decomposition, we propose that CAC should treat RA as a black box and exchange a minimal set of information. The questions remains as to what information will suffice. Does CAC need to know the power and subcarrier allocation? the resulting QoS? the utility earned as a function of time? We propose that CAC only needs to know  $\Lambda$ , the feasible region, and  $\{\bar{U}_{avg}(\mathbf{x}) \mid \forall \mathbf{x} \in \Lambda\}$ ,

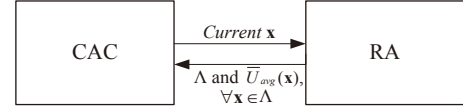


Fig. 1. Joint CAC and RA flowchart 1

the set of average utilities earned in each state by the resource allocation policy. Does RA need to know the history of arrivals and departures? We propose that RA only needs to know the current state. The resulting interface is illustrated in Fig 1.

Based on this interface, the decomposed CAC problem becomes one of choosing an admission policy that maximizes long-term average utility over the entire state space:

$$\max_{Q^{CAC}} \lim_{J \uparrow \infty} \frac{1}{t^{(J+1)}} \sum_{j=0}^J \bar{U}_{avg}(\mathbf{x}^{(j)}) \Delta t^{(j)} \quad (3)$$

In summary, the RA module determines the feasible region  $\Lambda$  and estimates the maximum average utility of each state  $\bar{U}_{avg}(\mathbf{x}^{(j)})$ , either by estimating utility as a function of the channel distribution or by direct observation of the real system. With this information, the CAC module can determine the optimal admission policy. Both of these roles can be done off-line. Then on-line, when a user arrives, the CAC module determines whether to admit merely using a lookup based on state, and the RA module determines power and subcarrier allocation based on the current state and the current set of channel gains.

Cooperation is thus instilled between CAC and RA by sharing the same optimization metric of long-term average utility. The CAC and RA modules can treat each other as black boxes, and communicate limited information via a simple interface. This decomposition provides flexibility to design CAC and RA policies separately and also reduces complexity.

## IV. IMPLEMENTATION

In this section, we give examples of RA and CAC algorithms that satisfy the decomposition and interface proposed in the previous section. This implementation will provide methods to estimate  $\bar{U}_{avg}(\mathbf{x})$  and to use dynamic programming to determine the optimal CAC policy.

### A. Resource Allocation

Whereas the model above assumed continuous time, in real systems resources are allocated during each time slot [10]. We thus would like to formulate a discrete time version of the RA problem. Define  $\Delta t_s$  as the duration of a time slot,  $t^l = \lceil t / \Delta t_s \rceil$ ,  $t^{(j),l} = \lceil t^{(j)} / \Delta t_s \rceil$  and  $[t^{(j),l}, t^{(j+1),l})$  as  $\{t^l \mid t^{(j),l} \leq t^l < t^{(j+1),l}\}$ . Then the set of time slots in which the system is in state  $\mathbf{x}$  is

$$T^l(\mathbf{x}) = \bigcup_{\{j \leq J: \mathbf{x}^{(j)} = \mathbf{x}\}} [t^{(j),l}, t^{(j+1),l})$$

and the discretized average utility is

$$U_{avg}^l(\mathbf{x}) = \lim_{J \uparrow \infty} \frac{1}{|T^l(\mathbf{x})|} \sum_{T^l(\mathbf{x})} \sum_{k \in D_{t^l}} U_{b_k}(S_{k,t^l})$$

where  $S_{k,t^\iota} = \sum_{\tau^\iota=t^\iota-W_k+1}^{t^\iota} R_{k,\tau^\iota} / W_k$ .

Denote the total number of active users while in state  $\mathbf{x}$  by  $K(\mathbf{x}) = \sum_{l=1}^L x_l$ . In different epochs  $j$  within  $T^\iota(\mathbf{x})$ , the total number of users is the same but the active user set  $D_{t^\iota}$  may be different. This makes it difficult to calculate  $\bar{U}_{avg}^\iota(\mathbf{x})$ . Thus for each state  $\mathbf{x}$ , we propose to consider a fixed set of users over an infinite length of time. For simplicity of notation, for each state  $\mathbf{x}$  we renumber the users from 1 to  $K(\mathbf{x}) = \sum_{l=1}^L x_l$  and restart time from 1 to  $T^\iota$ , and we therefore drop the  $\mathbf{x}$  from  $K(\mathbf{x})$ .

Denote the power allocation by  $\mathbf{p} = \{p_{k,n,t^\iota}, \forall k, n, t^\iota\}$ . Each subcarrier can be allocated to at most one user, thus denote the feasible set of power and subcarrier allocations by  $\mathbf{A} = \{\mathbf{p} \text{ s.t. } \forall t^\iota, n, p_{k,n,t^\iota} > 0 \text{ for at most one user } k\}$ . The discretized maximum average utility in state  $\mathbf{x}$  is then:

$$\begin{aligned} \bar{U}_{avg}^\iota(\mathbf{x}) &= \max_{\mathbf{p} \in \mathbf{A}} \frac{1}{T^\iota} \sum_{t^\iota=1}^{T^\iota} \sum_{k=1}^K U_{b_k}(S_{k,t^\iota}) \\ \text{s.t.} \quad &\sum_{k=1}^K \sum_{n=1}^N p_{k,n,t^\iota} \leq P \quad \forall t^\iota; \quad p_{k,n,t^\iota} \geq 0 \quad \forall k, n, t^\iota \\ &S_{k,t^\iota} > S'_{b_k} \quad \forall k, t^\iota \end{aligned} \quad (4)$$

We solve a dual problem by introducing a set of intermediate variables  $\mathbf{d} = \{d_{k,t^\iota}, \forall k, t^\iota\}$  as bounds on the achieved rates i.e.  $S_{k,t^\iota} \geq d_{k,t^\iota} \quad \forall k, t^\iota$ . The Lagrange is:

$$\begin{aligned} F(\mathbf{d}, \boldsymbol{\lambda}, \boldsymbol{\mu}) &= \frac{1}{T^\iota} \sum_{t^\iota=1}^{T^\iota} \sum_{k=1}^K U_{b_k}(d_{k,t^\iota}) + \sum_{t^\iota=1}^{T^\iota} \sum_{k=1}^K \lambda_{k,t^\iota} (S_{k,t^\iota} - d_{k,t^\iota}) \\ &\quad + \sum_{t^\iota=1}^{T^\iota} \mu_{t^\iota} \left( P - \sum_{k=1}^K \sum_{n=1}^N p_{k,n,t^\iota} \right) \end{aligned} \quad (5)$$

where  $d_{k,t^\iota} > S'_{b_k}$ ,  $\boldsymbol{\lambda} = \{\lambda_{k,t^\iota}, \forall k, 1 \leq t^\iota \leq T^\iota\}$  are the Lagrangian multipliers associated with the rate constraints which are interpreted as rate prices [6], and  $\boldsymbol{\mu} = \{\mu_{t^\iota}, 1 \leq t^\iota \leq T^\iota\}$  are the Lagrangian multipliers associated with the power constraints.

However, knowing the channel information for all  $1 \leq t \leq T^\iota$  is impractical and we propose to estimate  $\bar{U}_{avg}^\iota(\mathbf{x})$  using the channel distribution. Define a random variable  $\alpha_{k,n}^2$  representing fast fading for user  $k$  on subcarrier  $n$ , whose distribution is presumed known; denote  $\boldsymbol{\alpha}^2 = \{\alpha_{k,n}^2, \forall k, n\}$ . Combine slow fading, shadowing, and pathloss for user  $k$  into a single random variable  $\psi_k$ , whose distribution is presumed known; denote  $\boldsymbol{\psi} = \{\psi_k, \forall k\}$ . Denote the set of shadowing and pathlosses for all users except user  $k$  by  $\boldsymbol{\psi}_{-k} = \{\psi_{\hat{k}}, \forall \hat{k} \neq k\}$ . We assume that the distribution of combined slow fading and shadowing  $\psi_k$  is independent of  $k$ . As  $T^\iota \uparrow \infty$ ,

$$\frac{1}{T^\iota} \sum_{k=1}^K \sum_{t^\iota=1}^{T^\iota} U_{b_k}(S_{k,t^\iota}) \rightarrow \sum_{k=1}^K E_{\boldsymbol{\alpha}^2, \boldsymbol{\psi}} U_{b_k}(S_k)$$

Partition the domain of  $\psi_k$  into  $M$  slices, with the lower bound of slice  $m$  denoted by  $\psi^{(m)}$ . Denote the probability of

slice  $m$  by  $q_m = Pr(\psi_k = \psi^{(m)})$ . When in slice  $m$ , denote the power allocated to user  $k$  on subcarrier  $n$  by  $p_{k,n}^{(m)}$  and the average rate of user  $k$  by  $S_k^{(m)}$ . The optimal policy  $\mathbf{p}$  is then given by set of Lagrangian multipliers for each user denoted by  $\{\lambda_k^{(m)}, \forall k, m\}$ . The optimization problem (4) thus becomes:

$$\begin{aligned} \bar{U}_{avg}^\iota(\mathbf{x}) &\approx \max_{\{\lambda_k^{(m)}, \forall k, m\}} \sum_{k=1}^K \sum_{m=1}^M U_{b_k} \left( E_{\boldsymbol{\alpha}^2, \boldsymbol{\psi}_{-k}} S_k^{(m)} \right) q_m \\ \text{s.t.} \quad &\sum_{k=1}^K \sum_{m=1}^M \sum_{n=1}^N E_{\boldsymbol{\alpha}^2, \boldsymbol{\psi}_{-k}} \left( p_{k,n}^{(m)} \right) q_m \leq P \\ &S_k^{(m)} > S'_{b_k}, \forall k, m \end{aligned} \quad (6)$$

The detailed proof is in the Appendix. The discretized feasible region is

$$\Lambda^\iota = \left\{ \mathbf{x} : \exists \{\lambda_k^{(m)}, \forall k, m\} \text{ which can solve problem (6)} \right\}$$

### B. Connection Access Control

It remains to formulate and solve a discrete time version of problem (3). With the discretized feasible region, the CAC policy space becomes

$$G_{\mathbf{x}} = \{\mathbf{g} \in G : g_l = 0 \text{ if } \mathbf{x} + \mathbf{e}_l \notin \Lambda^\iota\}$$

where  $\mathbf{e}_l$  is a vector of zeros, except for a one for the  $l$ th component. The CAC policy  $Q^{CAC} = \mathbf{g}^{(j)}(\mathbf{x}^{(j-1)})$  remains to decide on admission of a connection on the basis of the state immediately before the connection's arrival.

The state  $\mathbf{x}^{(j)}$  is a continuous time Markov chain. It can be converted into a discrete time Markov chain via standard uniformization techniques [11](p110 and p209) as follows. The expected time in state  $\mathbf{x}^{(j)}$  under policy  $\mathbf{g}$  is

$$\nu_{\mathbf{x}}(\mathbf{g}) = \left[ \frac{1}{\sum_{l=1}^L \beta_l g_l + \sum_{l=1}^L x_l \gamma_l} \cdot \frac{1}{\Delta t_s} \right] \Delta t_s$$

and a bound on these expected times is

$$\nu = \left[ \frac{1}{\sum_{l=1}^L (\beta_l + \bar{x}_l \gamma_l)} \cdot \frac{1}{\Delta t_s} \right] \Delta t_s$$

where  $\bar{x}_l$  is the maximum number of class- $l$  applications in the network which can be calculated by allocating all the resources to type  $l$  applications. The embedded discrete time Markov chain transition probabilities are given by:

$$P_{\mathbf{x}\mathbf{x}'} = \begin{cases} x_l \gamma_l \nu_{\mathbf{x}}(\mathbf{g}), & l = 1, \dots, L, \text{ if } \mathbf{x}' = \mathbf{x} - \mathbf{e}_l \\ \beta_l g_l \nu_{\mathbf{x}}(\mathbf{g}), & l = 1, \dots, L, \text{ if } \mathbf{x}' = \mathbf{x} + \mathbf{e}_l \end{cases}$$

An equivalent discrete time problem is:

$$\max_{Q^{CAC}} \lim_{J \uparrow \infty} \frac{1}{J} \sum_{j=1}^J \bar{U}_{avg}^\iota(\mathbf{x}^{(j)}) \quad (7)$$

The optimal CAC policy can then be determined using stochastic dynamic programming. The corresponding value iteration algorithm [11](p210) is:

$$V_i(\mathbf{x}) = \max_{\mathbf{g} \in G_{\mathbf{x}}} \left\{ \bar{U}_{avg}^\iota(\mathbf{x}) + \frac{\nu}{\nu_{\mathbf{x}}(\mathbf{g})} \sum_{\mathbf{x}' \in \Lambda^\iota} P_{\mathbf{x}\mathbf{x}'}(\mathbf{g}) V_{i-1}(\mathbf{x}') \right\}$$

<sup>1</sup>We also presume that each user has been in the system since time  $t^\iota = -(W_{b_k} - 2)$ , so that  $S_{k,t^\iota}$  is defined at  $t^\iota = 1$ .

$$+ \left(1 - \frac{\nu}{\nu_{\mathbf{x}}(\mathbf{g})}\right) V_{i-1}(\mathbf{x}) \} \quad (8)$$

where  $i$  denotes the iteration number,  $V_i(\mathbf{x})$  denotes the expected return of state  $\mathbf{x}$ , and  $\mathbf{x}'$  is a possible next state.

In summary, the off-line portion of the CAC module uses stochastic dynamic programming to determine the optimal connection control policy; this is then used on-line to decide upon admissions. Upon admitted arrivals or departures, the CAC module passes the new state  $\mathbf{x}$  to the RA module. Off-line, the RA module uses problem (6) to determine a set of rate prices  $\{\lambda_k^{(m)} \forall k, m\}$ ; actually users in the same slice  $m$  using the same application  $l$  will face identical rate prices, denoted  $\lambda^{(l,m)}$ . These rate prices can be used on-line to allocate power and subcarriers based on the current state and the current set of channel gains using interpolation: if  $b_k = l$  and  $\psi^{(m)} < \psi_{k,t^e} < \psi^{(m-1)}$ , then an interpolated rate price can be set to:

$$\bar{\lambda}_{k,t^e} \approx \lambda^{(l,m)} - (\lambda^{(l,m)} - \lambda^{(l,m-1)}) \cdot \frac{1/\psi^{(m)} - \psi_{k,t^e}}{1/\psi^{(m)} - 1/\psi^{(m-1)}}$$

where  $\psi_{k,t^e}$  and  $\psi^{(m)}$  are expressed in dB.

## V. SIMULATION RESULT

In this section, we examine the performance of the proposed decomposition via simulation. We adopt the parameters of a LTE scenario [10]. The system bandwidth is 10MHz and total number of subcarriers is 1000. The base station transmission power is 46dBm with an antenna gain of 15 dBi. The inter-site distance is 750m. We assume 40% of the total resources are assigned to video users. Thus the total number of subcarriers for video users is  $N = 400$  and power constraint  $P = 42$ dbm. We set  $\sigma^2 + I$  equal to  $7.6232 \times 10^{-28}$  mW. All users move at a constant speed of 10km/h, with direction determined by a random walk [12]. The pathloss is:

$$PL_k = 128.1 + 37.6 \log_{10}(\zeta_k) + 21 \log_{10}(f_c/2.0) \text{ dB}$$

where  $\zeta_k$  is the distance from the user to the base station and  $f_c = 2000$ MHz is the central frequency. If users stay in the cell for a long period of time and/or if users' speed is high enough, these users are approximately uniformly distributed in a donut around the base station from  $\zeta = 0.01$ km to  $\zeta = 0.25$ km. The shadowing follows a lognormal distribution with mean value 0dB and variance 10dB [10]. The domain of  $\psi_k$  is partitioned into  $M = 35$  slices. Fast fading follows an Exponential distribution with mean value 1. The length of one time slot is 1 ms [13] and fast fading is generated every 3 time slots independent of previous fading.

We consider two kinds of semi-elastic applications and we set  $W_1 = W_2 = 133$  slots which is the length of one group of pictures in MPEG4. The utility function for a type  $l$  application is:

$$U_l(S_{k,t^e}) = \begin{cases} a_l(S_{k,t^e})^2, & \text{if } S_{k,t^e} < S_l^f \\ c_l(S_{k,t^e} + b_l)^{1/3}, & \text{else} \end{cases}$$

where  $S_{k,t^e}$  is also expressed in units of 100kbps,  $a_1 = 4/5 * (2/5)^{1/3}/(12/5)^2$ ,  $b_1 = -2$ ,  $c_1 = 4/5$ ,  $S_1^f = 240$ kbps,  $a_2 =$

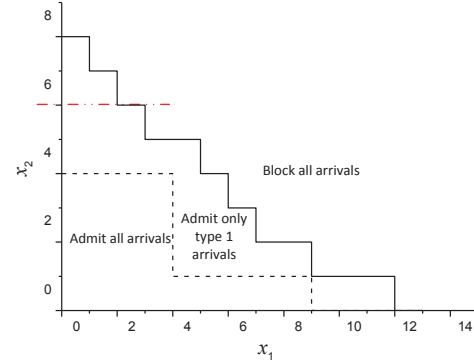


Fig. 2. State Space and Admission Policy

TABLE I  
AVERAGE UTILITY

	Traditional CAC	Proposed CAC
Utility	25.89	32.49

TABLE II  
MINIMUM AVERAGE RATE (IN UNITS OF KBPS)

Slice number	Traditional CAC		Proposed CAC	
	avg $S_1^m$	avg $S_2^m$	avg $S_1^m$	avg $S_2^m$
$m = 32$	494	756	529	753
$m = 33$	498	696	489	727
$m = 34$	465	661	476	729
$m = 35$	429	641	479	635

$(5/6)^{1/3}/25$ ,  $b_2 = -25/6$ ,  $c_2 = 1$ ,  $S_2^f = 500$ kbps. Both utility functions are sigmoid [6]; we set the minimum rate requirements to the rate at the maximum average utility, which is given by  $S_1' = 300$ kbps and  $S_2' = 625$ kbps. The arrival processes are Poisson processes with intensities  $\beta_1 = 1/12/s$  and  $\beta_2 = 1/30/s$  respectively, and the connection durations are Exponentially distributed with means  $1/\gamma_1 = 180s$  and  $1/\gamma_2 = 300s$  respectively.

The state space and admission policy are shown in Fig 2. The cell can admit  $\bar{x}_1 = 11$  type 1 applications,  $\bar{x}_2 = 7$  type 2 applications, or a bit less than a linear combination of the two. In general, type 1 applications are preferred, since they generate a higher average utility per unit rate at their minimum rate, i.e.  $U_1(S_1')/S_1' > U_2(S_2')/S_2'$ . When the state is on or above the upper boundary of the state space, the optimal CAC policy is to block all arrivals. When the state is far from the upper boundary, the optimal policy is to admit all arrivals. In a zone near the boundary as illustrated, the optimal policy is to admit only type 1 arrivals, i.e. arrivals of type 1 applications; as a result, the system will never enter a state with  $x_2 > 5$ .

We compare our coordinated admission policy with a traditional admission policy which admits arriving users if and only if there are enough resources to guarantee QoS constraints, i.e. one that admits all users in all states under the upper boundary. The same RA module in section IV is used for both of these two policies. We simulate 40 minutes of arrivals and departures; the resulting average utility under each policy is shown in Table I. In this situation, the coordinated algorithm achieves a much higher average utility than the traditional uncoordinated algorithm. This occurs because the coordinated

CAC policy results in the system spending a substantially greater portion of time near the highest utility state (11,0) than does the traditional policy.

Finally, we examine how well each CAC and RA policy adheres to their given QoS constraints. To do this, we focus on the outermost 4 slices,  $m = 32$  to 35, which experience the lowest average rates. In Table II, we show the minimum average rate of all users in the given slice within each application type. We observe that both the coordinated and uncoordinated CAC and RA policies maintain average rates above the QoS requirement  $S_i$  for all users.

## VI. CONCLUSION

We have investigated how connection access control and resource allocation can be coordinated so that both focus on utility. Whereas previous CAC proposals have typically admit a new connection if and only if it is believed that capacity is available to ensure acceptable performance, here we have proposed that CAC and RA should both focus on maximizing long-term average utility.

We have found that joint CAC and RA can be decomposed into separate CAC and RA modules that treat each other as black boxes, and that the interface simply consists of an exchange of information about feasible states, current state, and average utilities in each state. This allows the RA module to maximize the long-term average utility for each system state, and the CAC module to maximize the long-term average utility over all system states. This decomposition provides flexibility in the design of the CAC and RA policies.

We also gave examples of RA and CAC algorithms that satisfy this decomposition and interface. The RA module determines a set of rate prices for each application based on a user's combined slow fading, shadowing and pathloss, and uses these rate prices to allocate power and subcarriers. The CAC module uses stochastic dynamic programming to determine the optimal admission policy.

Simulation showed that the optimal CAC policy may block applications with relatively low average utility per unit rate even when capacity is available. This preferential treatment by application can result in much higher average utility while continuing to satisfy QoS constraints.

## REFERENCES

- [1] C. Y. Wong, R. Cheng, K. Lataief, and R. Murch, "Multiuser OFDM with adaptive subcarrier, bit, and power allocation," *IEEE J. Sel. Areas Commun.*, vol. 17, no. 10, pp. 1747–1758, Oct. 1999.
- [2] P. Liu, P. Zhang, S. Jordan, and M. Honig, "Single-cell forward link power allocation using pricing in wireless networks," *IEEE Trans. Wireless Commun.*, vol. 3, no. 2, pp. 533–543, Mar. 2004.
- [3] J. Jang and K. B. Lee, "Transmit power adaptation for multiuser OFDM systems," *IEEE J. Sel. Areas Commun.*, vol. 21, no. 2, pp. 171–178, Feb. 2003.
- [4] G. Song and Y. Li, "Utility-based resource allocation and scheduling in OFDM-based wireless broadband networks," *IEEE Commun. Mag.*, vol. 43, no. 12, pp. 127–134, Dec. 2005.
- [5] C. Yang and S. Jordan, "Resource allocation for semi-elastic applications in wireless networks," in *Global Telecommunications Conference (GLOBECOM 2011)*, 2011 IEEE, Dec. 2011, pp. 1–6.
- [6] —, "Power and rate allocation for video conferencing in cellular networks," in *Communication, Control, and Computing (Allerton)*, 2011 49th Annual Allerton Conference on, Sept. 2011, pp. 127–134.

- [7] Z. Liu and M. El Zarki, "SIR-based call admission control for DS-CDMA cellular systems," *Selected Areas in Communications, IEEE Journal on*, vol. 12, no. 4, pp. 638–644, May 1994.
- [8] S. Bashar and Z. Ding, "Admission control and resource allocation in a heterogeneous OFDMA wireless network," *Wireless Communications, IEEE Transactions on*, vol. 8, no. 8, pp. 4200–4210, August 2009.
- [9] X. Gong, S. Vorobyov, and C. Tellambura, "Joint bandwidth and power allocation with admission control in wireless multi-user networks with and without relaying," *Signal Processing, IEEE Transactions on*, vol. 59, no. 4, pp. 1801–1813, April 2011.
- [10] *E-UTRA Radio Frequency (RF) system scenarios*, 3GPP TR 36.942 v10.2.0 Std.
- [11] H. C. Tijms, *Stochastic Modelling and Analysis: A Computational Approach*. New York: Wiley, 1986.
- [12] *Selection procedures for the choice of radio transmission technologies of the UMTS*, 3GPP TR 30.03 Std.
- [13] *Evolved Universal Terrestrial Radio Access*, 3GPP TR 36.300 v10.2.0 Std.

## VII. APPENDIX

The first order conditions for (5) result in an optimal power allocation:

$$p_{k,n,t^t} = \left( \frac{B\bar{\lambda}_{k,t^t}}{\mu_{t^t} \ln 2} - \frac{\sigma^2 + I}{|H_{k,n,t^t}|^2} \right)^+ \quad (9)$$

and that subcarrier  $n$  should be allocated to the user

$$\arg \max_k \Phi_{k,n,t^t} \quad (10)$$

where

$$\Phi_{k,n,t^t} = \bar{\lambda}_{k,t^t} B \left[ \log_2 \left( \frac{B\bar{\lambda}_{k,t^t}}{\mu_{t^t} \ln 2} \frac{|H_{k,n,t^t}|^2}{\sigma^2 + I} \right) \right]^+ - \mu_{t^t} \left( \frac{B\bar{\lambda}_{k,t^t}}{\mu_{t^t} \ln 2} - \frac{\sigma^2 + I}{|H_{k,n,t^t}|^2} \right)^+$$

where  $\bar{\lambda}_{k,t^t} = \sum_{\tau^t=t^t}^{t^t+W-1} \lambda_{k,\tau^t} / W_k$ . The target rate is

$$d_{k,t^t} = \max \left\{ S'_{b_k}, \arg \max_{d_{k,t^t}} [U_{b_k}(d_{k,t^t}) / T^t - \lambda_{k,t^t} d_{k,t^t}] \right\} \quad (11)$$

We propose basing the average rate price  $\bar{\lambda}_{k,t^t}$  on a user's combined pathloss and shadowing which is nearly constant during time window  $W_{b_k}$ . The fluctuations in fast fading will average out during time window  $W_{b_k}$ , whereas fluctuations in slow fading, shadowing, and pathloss will not. Thus the rate price is principally influenced by combined path loss and shadowing, i.e.  $\bar{\lambda}_{k,t^t}$  should be a function of  $\psi_k$ , denoted  $\lambda_k(\psi_k)$ . User  $k$ 's achieved rate  $S_k$  depends not only on its  $\psi_k$  but also upon other users'  $\psi_{-k}$  and on all users' fast fading. Since fast fading will largely average out within time window  $W_{b_k}$ , the expectation over  $\alpha^2$  can be brought inside the utility function without loss of accuracy. Other users' shadowing and pathlosses affect the resource allocation through determination of the power price  $\mu_{k,t^t}$ , but consideration of them in determination of the resource allocation policy  $\{\lambda_k(\psi_k), \forall k\}$  is too complex; thus we also bring  $\psi_{-k}$  inside the utility function. Then

$$\sum_{k=1}^K E_{\alpha^2, \psi} U_{b_k}(S_k) \approx \sum_{k=1}^K E_{\psi_k} U_k(E_{\alpha^2, \psi_{-k}} S_k)$$

We partition  $\psi_k$  into  $M$  slices. This leads to the quantized optimization problem (6).