Proceedings of the
44th IEEE Conference on Decision and Control, and
the European Control Conference 2005
Seville, Spain, December 12-15, 2005

Jointly Optimal MAC and Transport Layers in CDMA Broadband Networks

Jennifer Price and Tara Javidi
Department of Electrical and Computer Engineering
University of California, San Diego
La Jolla, CA 92093
jenn-price@ucsd.edu, tara@ece.ucsd.edu

Abstract— **In this paper, we focus on cross-layer (MAC and transport) design for variable-rate CDMA networks. First, we formulate the cross-layer rate assignment task as a constrained convex optimization problem. Next, we develop two sets of distributed feedback algorithms to solve this problem - one which merges the rate assignments at the MAC and transport layers (one-shot), and one which coordinates them (modular). We show that both sets of algorithms converge to the same equilibrium. In particular, we show that the addition of queues between the transport and MAC layers can actually facilitate the coordination required for the modular algorithm with only minimal modification to existing protocols. Practical distributed implementation and its impact on the convergence of both algorithms is addressed.**

I. Introduction

With the advent of 3G and 4G networks, the challenges of optimal resource allocation have moved to the forefront of wireless network design. Not only is the wireless medium a shared and limited resource, but it suffers from randomly fluctuating channel conditions. In a data-only network where users can tolerate variable transmission rates and delays, we would like to respond to these random fluctuations by adapting transmission rates for efficient channel utilization. This is similar to that of modern IP protocols for wired networks. Unlike users in a wired network, however, a wireless user's rate is regulated by both the transport layer (in response to the congestion status of the links in the core), and the MAC layer (in response to the interference levels and channel quality of the wireless medium). Traditional network design favors a "black-box" approach in which the transport and MAC layer protocols are designed and implemented separately, but this approach inherently limits the information available to each module and may prohibit an optimal resource allocation. In addition, interactions between transport and MAC layers can significantly degrade performance in terms of both throughput and delay [2], [17]. The recognition of the fact that the "black box" approach may not be optimal for wireless networks has led to a renewed interest in cross-layer design.

In this paper, we consider a wideband CDMA network with variable transmission rates, similar to the CDMA2000 and 1xEVDO systems. Mobiles communicate with the base stations (single-hop), which are connected directly to the wired IP network. In other words, we have a multi-hop

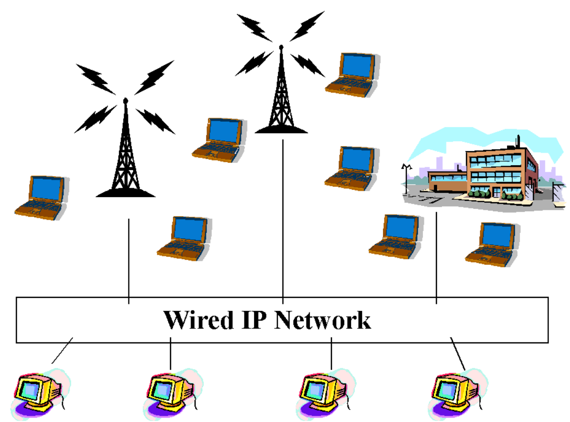

Fig. 1. Single-Hop Cellular Network Structure

network in which the first hop for some users is a wireless CDMA link, as shown in Figure 1. Using this context, we formulate the optimal rate assignment task as a constrained convex optimization problem. We then develop two sets of distributed feedback algorithms to solve this problem. The first set of algorithms algorithm merges the rate control functionality of the MAC and transport layers by assigning a single rate for each user which is a function of both interference and congestion constraints, as shown in Figure 2(a). This is referred to as the *one-shot* approach. The problem with this approach is, as with many cross-layer designs, that it does not respect the modularity of the protocol stack. With this in mind, we develop a second set of algorithms which allows for a *coordinated* control of rate at both the MAC and transport layers. This approach, referred to as the *modular* approach, allows the MAC and transport protocols to be implemented separately, as shown in Figure 2(b). We then show that both the one-shot and modular algorithms result in the same (cross-layer) optimal rate allocation. Furthermore, we show that the modular algorithm can actually be implemented with only minimal modification to existing protocols.

The remainder of the paper is organized as follows. Section II contains background material including related works, network setting and CDMA interference models, and the formulation of the rate-assignment task as a constrained optimization problem. Section III details the design of the

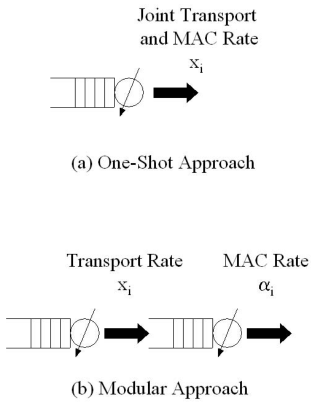

Fig. 2. Internal Structure of a Wireless User

one-shot rate assignment algorithm, while Section IV details the design of the modular algorithm. Section V describes how to implement these algorithms in a distributed manner, and addresses the impact that the distributed implementations have on the convergence of the algorithms. Finally, Section VI discusses possible performance trade-offs between the two algorithms, and details areas for further investigation.

II. Background Material

A. Related Works

We choose to examine the resource allocation problem using an optimization framework. In other words, we seek to optimize the rate assignments subject to the interference and congestion constraints dictated by the wireless and wired mediums. Conceptually, this is similar to a large body of work on congestion control in wired TCP (see [10], [12], [15], [16], [25]). This work on congestion control in wired networks has been extended to wireless networks in various contexts, addressing such issues as multiple QOS constraints, cross-layer design, rate control, power control, or capacity constraints (see [4], [5], [7]-[9], [18]-[24]).

In this, our work is somewhat related to the cross-layer design in [5]. In this work, the authors address joint rate and power control for ad-hoc wireless networks. The main difference between this and our work is the treatment of wireless constraint. The authors in [5] treat wireless connections as links with variable capacity (using a generalized notion of information theoretic capacity). This allows the authors to decompose the problem into separate rate and power assignments using message passing. Such a characterization of capacity is problematic, since a time-varying capacity need not be a sufficient constraint on decodability. This issue becomes even more critical in the context of delay. On the other hand, we work directly with a constraint on SINR, which is far more practical from an implementation standpoint. Such a constraint does not allow for the decomposition used in [5], making the practical distributed implementation of our scheme more challenging.

In addition, our work is related to the cross-layer design in [14]. In this work, the authors address joint scheduling and rate control in wireless multi-hop networks. The idea of introducing queues at the wireless link in these papers is similar to our work. The introduction of these queues (which we refer to as MAC layer queues) allows for the overlay of two sets of rate control problems: 1) at the transport layer, and 2) at the wireless link. The main difference between [14] and our work is in the formulation of the latter problem. In [14], this problem is addressed as a general time scheduling rate-control problem. This hinders the development of distribution solutions at the wireless link. On the other hand, we restrict our attention to a CDMA system, where we can use the dynamic variations of the spreading gain, feasible rate region, and bases' signaling to provide optimal rate control at the wireless link.

B. Network Setting and CDMA Interference Model

We use the following notation. There are a total of M sources transmitting with transport-layer rate x_i. Without loss of generality, the set of all sources can be ordered as $\{1, \ldots, N, N+1, \ldots, M\}$, where the first N are wireless sources. Each wireless source transmits over the air with MAC-layer rate α_i.

There is a set $\mathcal{J} = \{1, 2, \ldots, N\}$ of wired links, each with capacity C_j. The set of wired links used by source i is fixed, and denoted by l_i. The routing function is defined as

$$\psi_{ij} = \begin{cases} 1 & \text{if} j \in l_i \\ 0 & \text{if} j \notin l_i \end{cases}$$

There is a set $\mathcal{L}$ of L CDMA-based wireless sectors associated with wireless access points (bases). The tracking base for wireless source i, denoted $b(i)$, is the base to which wireless source i is connected. This is also the base responsible for wireless source i's power control. For simplicity, we assume that each wireless source is tracked by exactly one base, and each base tracks exactly one sector. P_i is the transmitted power for user i, and g_{il} is the channel gain (assumed to be fixed). W is the chip bandwidth, and N_0 is the thermal noise density. The spreading gain for mobile i is defined as $s_i = \frac{W}{\alpha_i}$.

Consider wireless source i which is tracked by sector $l = b(i)$. The bit energy per interference power density of mobile i at base station l can be written as

$$\frac{E_b}{I_0}^l(i) = \frac{s_i P_i g_{il}}{N_0 W + \sum_{k=1, k \neq i}^{N} P_k g_{kl}} \qquad (1)$$

where N_0 is the thermal noise density. The signal-to-noise ratio of mobile i at base station l can then be written as $SINR^l(i) = \frac{E_b}{I_0}^l(i)(\frac{\alpha_i}{W})$.

Notice that in an interference limited system such as CDMA, the relationship between SINR and information rate given by the Shannon equation can be approximated as a linear one (i.e. $\log(1+y) \approx y$ when $y \ll 1$) [13]. This means that an increase in MAC rate α_i translates directly into a linear increase in information rate if and only if $\frac{E_b}{I_0}$ is kept the same (e.g. at γ). In other words, we assume the condition $\frac{E_b}{I_0} = \gamma$ is a necessary condition for decodability of transmissions of information at a rate proportional to α_i [26].

C. Cross-Layer Optimization Problem

In order to address rate-control as a constrained optimization problem, we must first introduce the notion of feasible rate assignments. We say a pair of rate vectors $(x_1, \ldots, x_M)$ and $(\alpha_1, \ldots, \alpha_N)$ is a feasible rate solution if there exists a power vector $(P_1, \ldots, P_N)$ such that the following conditions are satisfied:

- $\sum_{i=1}^{M} x_i \psi_{ij} \leqslant C_j \quad \forall j \in \mathcal{J}$
- $\sum_{i=1}^{N} P_i g_{il} \leqslant K N_0 W \quad \forall l \in \mathcal{L}$
- $\frac{E_b}{I_0}^l(i) = \gamma \quad \forall i \leqslant N$ and $l = b(i)$
- $x_i = \alpha_i \quad \forall\, i \leqslant N$

where γ is a pre-specified value (see previous section for further discussion).

In order to establish the feasibility of a vector of rates, we need to first solve Eqn (1) to calculate the appropriate power vector, then establish the validity of the above conditions. We see that the first condition is simply the link capacity constraint for the wired network [15], and depends on the routing matrix ψ. The second condition is used as an alternative to limiting individual transmission power at each wireless source [1], and depends on both sector assignments and the channel conditions. The third condition guarantees an acceptable BER for wireless transmissions, and the fourth condition guarantees the stability of MAC layer queues.

In [18] and [21], we have developed a simpler feasibility region with a linear-type structure. Using similar modifications, we formally define our feasibility region.

Definition 1: A pair of rate vectors $(x_1, \ldots, x_M)$ and $(\alpha_1, \ldots, \alpha_N)$ belongs to the *feasible region* Δ if and only if it satisfies the following conditions:

C1. $\sum_{i=1}^{M} x_i \psi_{ij} \leqslant C_j \; \forall\, j \in \mathcal{J}$

C2. $\sum_{i=1}^{N} \frac{\alpha_i}{W + \gamma\alpha_i} \frac{g_{il}}{g_{ib(i)}} \leqslant \frac{K}{\gamma(1+K)} \; \forall\, l \in \mathcal{L}$

C3. $x_i = \alpha_i \quad \forall\, i \leqslant N$

Having established a region of feasible rate vectors, we wish to choose a rate vector that is proportional fair [10]. This is equivalent to optimizing the utility function $\sum_{i=1}^{M} \log(x_i)$, resulting in the following optimization problem:

P. Find the pair of rate vectors $(\underline{x}, \underline{\alpha})$ that is the solution to:

$$(\underline{x}^*, \underline{\alpha}^*) = \arg \max_{(\underline{x}, \underline{\alpha}) \in \Delta} \sum_{i=1}^{M} \log(x_i)$$

In the remainder of the paper, we will see how optimization and dual theory can be used to develop distributed solutions to Problem P.

III. Distributed Algorithm I: One-Shot Rate Assignment

We begin our discussion of distributed rate-assignment algorithms with a one-shot approach which follows the structure of Figure 2 (b). Rather than having separate MAC and transport layer rates, we work with a single rate x_i by absorbing Condition C3 (i.e. replacing α_i with x_i in Condition C2). This results in a constrained convex optimization problem whose associated Lagrangian is:

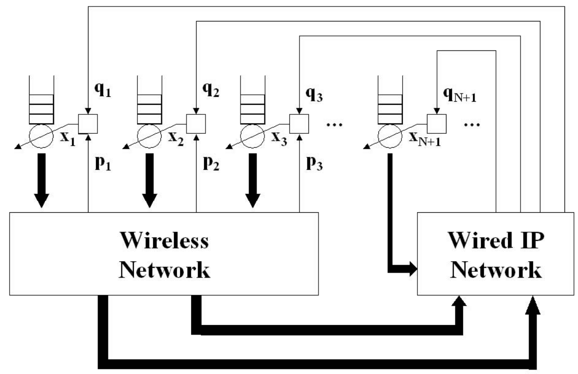

Fig. 3. Control-Theoretic View of One-Shot Rate Assignments

$$\mathcal{L}_C(\underline{x}, \underline{\lambda}, \underline{\mu}) = \sum_{i=1}^{M} \log(x_i) - \sum_{j=1}^{J} \lambda_j \left(\sum_{i=1}^{M} x_i \psi_{ij} - C_j \right) - \sum_{l=1}^{L} \mu_l \left(\sum_{i=1}^{N} \frac{g_{il}}{g_{ib(i)}} \frac{x_i}{W + \gamma x_i} - \frac{K}{\gamma(1+K)} \right)$$

The dual problem can then be formulated as follows:

DP. Find the Lagrange multipliers $(\lambda_1, \ldots, \lambda_J)$ and $(\mu_1, \ldots, \mu_L)$ such that they solve

$$\min_{\underline{\mu}, \underline{\lambda} \geqslant 0} \sum_{i=1}^{M} \phi_i(q_i, p_i) + \sum_{j=1}^{J} \lambda_j C_j + \frac{K}{\gamma(1+K)} \sum_{l=1}^{L} \mu_l$$

where

$$q_i = \sum_{j=1}^{J} \lambda_j \psi_{ij}$$

$$p_i = \begin{cases} \sum_{l=1}^{L} \frac{g_{il}}{g_{ib(i)}} \mu_l & i = 1, \ldots, N \\ 0 & i = N+1, \ldots, M \end{cases}$$

$$\phi_i(q_i, p_i) = \max_{x} \left(\log(x) - x q_i - \frac{x}{W + \gamma x} p_i \right)$$

Notice that for a given set of Lagrange multipliers, $\phi_i(q_i, p_i)$ is an autonomous rule that can be implemented at each source using locally available information. This is an extremely attractive property since it allows for distributed computation of the rate assignments. When the multipliers are chosen appropriately, the autonomous rule $\phi_i(q_i, p_i)$ results in a globally optimal, proportional fair solution. If we continually update these multipliers, we can actually respond to changes in the network by constructing a distributed feedback loop, as shown in Figure 3. This can be done by using a gradient projection method to generate the multipliers. The wired and wireless networks generate regulatory signals through gradient projection. These signals combine to form aggregate signals which, in turn, are used by the sources to select their transmission rate. This not only facilitates distributed computation of the rate assignments, but also allows for continuous adaptation to changing network conditions. The resulting algorithm consists of the following three components:

Base Algorithm

Each base station produces a regulatory signal (Lagrangian multiplier μ_l) that indicates the level of interference at that sector. This signal evolves according to the following difference equation:

$$\Delta\mu_l = \begin{cases} \beta(\sum_{i=1}^{N} \frac{g_{il}}{g_{ib(i)}} \frac{x_i}{W+\gamma x_i} - \frac{K}{\gamma(1+K)}) & \text{if } \mu_l(t) > 0 \\ \beta[\sum_{i=1}^{N} \frac{g_{il}}{g_{ib(i)}} \frac{x_i}{W+\gamma x_i} - \frac{K}{\gamma(1+K)}]^{+} & \text{if } \mu_l(t) = 0 \end{cases} \tag{2}$$

where β is a constant and $\sum_{i=1}^{N} \frac{g_{il}}{g_{ib(i)}} \frac{x_i}{W+\gamma x_i}$ is a measure of interference at each sector. These signals are then used to generate the aggregate signals $\underline{p}$.

Wired Link Algorithm

Each link produces a regulatory signal (Lagrangian multiplier λ_j) that indicates the level of congestion at that link. This signal evolves according to the following difference equation:

$$\Delta\lambda_j = \begin{cases} \xi(\sum_{i=1}^{M} x_i\psi_{ij} - C_j) & \text{if } \lambda_j(t) > 0 \\ \xi[\sum_{i=1}^{M} x_i\psi_{ij} - C_j]^{+} & \text{if } \lambda_j(t) = 0 \end{cases} \tag{3}$$

where ξ is a constant and $\sum_{i=1}^{M} x_i\psi_{ij}$ is the total traffic on link j. These signals are then used to generate the aggregate signals $\underline{q}$.

Source Algorithm

Each source reacts to the levels of interference and congestion, indicated by the base station and link coordination signals, by adjusting its rate such that

$$x_i = \arg\max_x \left(\log(x) - xq_i - \frac{x}{W+\gamma x}p_i\right) \tag{4}$$

IV. Distributed Algorithm II: Modular Rate Assignments

The previous section describes a distributed one-shot rate assignment algorithm which produces a transmission rate at each source. This requires a complete elimination of the protocol stack for wireless sources. In reality, it is desirable to implement transport and MAC layer protocols in separate modules following the structure of Figure 2 (a). In such a case, each wireless user has both a transport layer rate x_i, and a MAC layer rate α_i. Throughout this section, we seek to develop a distributed rate assignment algorithm with this structure that still converges to the same (cross-layer) optimal point, i.e. the solution to Problem P.

The challenge in directly applying dual methods to Problem P is that $\sum_{i=1}^{M} \log(x_i)$ is not concave in MAC rates α_i. In order to remedy this we write the utility of wireless user i as:

$$\begin{aligned} \log(x_i) &= (1-\sigma)\log(x_i) + \sigma\log(x_i) \\ &= (1-\sigma)\log(x_i) + \sigma\log(\alpha_i) \end{aligned}$$

where $0 < \sigma < 1$ is a constant. This does not change the solution to the problem since Condition C3 ensures $x_i = \alpha_i$. To simplify notations, we define

$$V_i(x_i) = \begin{cases} \log(x_i) & \text{if } i > N \\ (1-\sigma)\log(x_i) & \text{if } i \leqslant N \end{cases}$$

We now have the following modified problem statement:

P'. Find the pair of rate vectors $(\underline{x}, \underline{\alpha})$ that is the solution to:

$$(\underline{x}^*, \underline{\alpha}^*) = \arg\max_{\underline{x},\underline{\alpha}\in\Delta} \sum_{i=1}^{N} \sigma\log(\alpha_i) + \sum_{i=1}^{M} V_i(x_i)$$

As with Problem P, we again have a constrained convex optimization problem. Consider the Lagrangian associated with P':

$$\begin{aligned} L_M(\underline{x}, \underline{\alpha}, \underline{\lambda}, \underline{\mu}, \underline{\nu^+}, \underline{\nu^-}) &= \sum_{i=1}^{N}(1-\sigma)\log(\alpha_i) + \sum_{i=1}^{M} V_i(x_i) \\ &- \sum_{j=1}^{J} \lambda_j \left(\sum_{i=1}^{M} x_i\psi_{ij} - C_j\right) \\ &- \sum_{l=1}^{L} \mu_l \left(\sum_{i=1}^{N} \frac{g_{il}}{g_{ib(i)}} \frac{\alpha_i}{W+\gamma\alpha_i} - \frac{K}{\gamma(1+K)}\right) \\ &- \sum_{i=1}^{N} \nu_i^+(x_i - \alpha_i) - \sum_{i=1}^{N} \nu_i^-(\alpha_i - x_i) \end{aligned}$$

The dual problem can then be formulated as follows:

DP'. Find the Lagrangian multipliers $(\lambda_1, \ldots, \lambda_J)$, $(\mu_1, \ldots, \mu_L)$, $(\nu_1^+, \ldots, \nu_N^+)$ and $(\nu_1^-, \ldots, \nu_N^-)$ such that they solve

$$\min_{\underline{\lambda},\underline{\mu},\underline{\nu}\geqslant 0} \quad \sum_{i=1}^{M} \phi_i(q_i, \nu_i) + \sum_{i=1}^{N} \rho_i(p_i, \nu_i) + \sum_{j=1}^{J} \lambda_j C_j + \frac{K}{\gamma(1+K)} \sum_{l=1}^{L} \mu_l$$

where q_i and p_i are as defined in Section III, and

$$\nu_i = \begin{cases} \nu_i^+ - \nu_i^- & \text{if } i \leqslant N \\ 0 & \text{if } i > N \end{cases}$$

$$\phi_i(q_i, \nu_i) = \max_x (V_i(x) - x(q_i + \nu_i))$$

$$\rho_i(p_i, \nu_i) = \max_\alpha \left((1-\sigma)\log(\alpha) + \alpha\nu_i - \frac{\alpha}{W+\gamma\alpha}p_i \right)$$

Similar to the one-shot design, $\phi_i(q_i, \nu_i)$ and $\rho_i(p_i, \nu_i)$ are autonomous rules that can be implemented at each source using locally available information. In this case, however, we see the addition of a "cross-layer coordination signal", ν_i. This signal is used to coordinate each user's two separate rate adjustments: one at the transport layer (x_i), and one at the MAC layer (α_i). Again we use gradient projection to generate the Lagrange multipliers. This allows us to construct a distributed feedback loop that continually adapts to changing network conditions, shown in Figure 4. Although this is the same basic structure as the one-shot design, we now have an inner feedback loop corresponding to the MAC layer rate assignment and an outer feedback loop corresponding to the

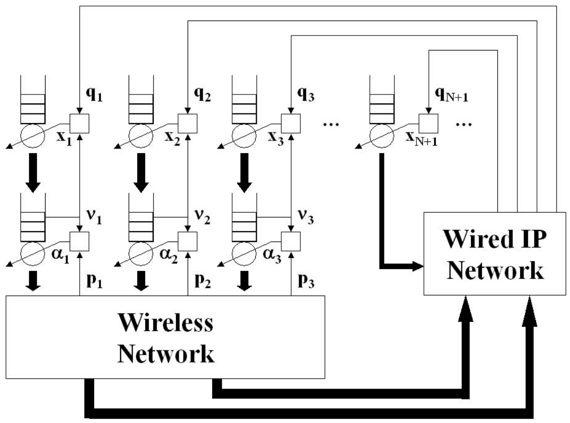

Fig. 4. Control-Theoretic View of Modular Rate Assignments

transport layer rate assignment. The cross-layer coordination signal is used to coordinate the operation of these two loops. The resulting algorithm consists of five components:

Base Algorithm

Identical to the Base Algorithm presented in Section III, except that x_i is replaced by α_i in Eqn (2).

Link Algorithm

Identical to the Link Algorithm presented in Section III.

Wireless Algorithm

Each wireless source produces two internal coordination signals (Lagrangian multipliers ν_i^+ and ν_i^-) that indicate the difference between transport and MAC layer rates at that source. These signals evolve according to the following difference equations:

$$\Delta\nu_i^+ = \begin{cases} \zeta_1(x_i - \alpha_i) & \text{if } \nu_i^+(t) > 0 \\ \zeta_1[x_i - \alpha_i]^+ & \text{if } \nu_i^+(t) = 0 \end{cases} \quad (5)$$

and

$$\Delta\nu_i^- = \begin{cases} \zeta_2(\alpha_i - x_i) & \text{if } \nu_i^-(t) > 0 \\ \zeta_2[\alpha_i - x_i]^+ & \text{if } \nu_i^-(t) = 0 \end{cases} \quad (6)$$

where ζ is a scalar. Note that these signals are generated only at the wireless sources, and are then used to generate the aggregate signals $\underline{\nu}$.

Transport Layer Source Algorithm

Each source reacts to the levels of congestion (indicated by the link coordination signals) and the mismatch between transport and MAC layer rates (indicated by the cross-layer coordination signals) by adjusting its transport-layer rate such that

$$x_i = \arg\max_x (V_i(x) - x(q_i + \nu_i)) \quad (7)$$

This algorithm is run at both wired and wireless sources.

MAC Layer Source Algorithm

Each wireless source reacts to the interference levels at each sector (indicated by the base coordination signals) and the mismatch between transport and MAC layer rates (indicated by the cross-layer coordination signals) by adjusting its MAC-layer rate such that

$$\alpha_i = \arg\max_\alpha (\sigma \log(\alpha) + \alpha\nu_i - \frac{\alpha}{W + \gamma\alpha} p_i) \quad (8)$$

This algorithm is run only at the output of wireless sources.

V. Practical Distributed Implementation

Although the formulations described in Sections III and IV allow for parallel computations, this does not necessarily correspond to a practical distributed control mechanism. In other words, the Lagrange multipliers need not be locally available, even though they can always be computed in parallel. Previously, we have addressed the practical implementation of the one-shot algorithm (see [20]). In this section, we show that the modular algorithm can also be implemented in a distributed manner with reasonable overhead. We also discuss how the distributed implementation impacts the convergence results for both sets of algorithms.

A. Signaling Mechanisms

Since the computation and communication of regulating signals is the basis of the distributed algorithms described in previous sections, it is natural to start by discussing how this information is exchanged in a practical setting. In particular, we are interested in addressing:

1. the computation of the regulating signals $\underline{\mu}$ and the availability of the corresponding aggregate signals $\underline{p}$
2. the computation of the regulating signals $\underline{\lambda}$ and the availability of the corresponding aggregate signals $\underline{q}$
3. the computation of the regulating signals $\underline{\nu^+}$ and $\underline{\nu^-}$.

Recall the base algorithm from Eqn (2). This equation requires each base to know information about the load at all other bases. In order to facilitate distributed computation, we introduce the following alternative which *approximates* the original solution:

$$\Delta\mu_l \backsimeq \begin{cases} \beta(\sum_{i=1}^N \frac{P_i g_{il}}{N_0 W} - K) & \text{if } \mu_l(t) > 0 \\ \beta[\sum_{i=1}^N \frac{P_i g_{il}}{N_0 W} - K]^+ & \text{if } \mu_l(t) = 0 \end{cases} \quad (9)$$

The quantity $\sum_{i=1}^N \frac{P_i g_{il}}{N_0 W}$ can be measured at each base station [1], and represents the overall interference. We refer to this quantity as *Rise Over Thermal* (ROT).

Once the regulating signals μ_l are computed at each base, they are used to generate aggregate signals for each mobile. Recall the definition of each user's aggregate wireless signal $p_i = \sum_{l=1}^L \frac{g_{il}}{g_{ib(i)}}$. At first glance it seems that in order to calculate p_i, each mobile requires full knowledge of the channel. In [18] and [21] we have shown that there exists a practical solution to this problem using the CDMA pilot signal, PS, and a pricing pilot signal, PPS. This pilot symbol is transmitted with a power level proportional to the base signal, μ_l. Hence p_i can be calculated as $p_i =$

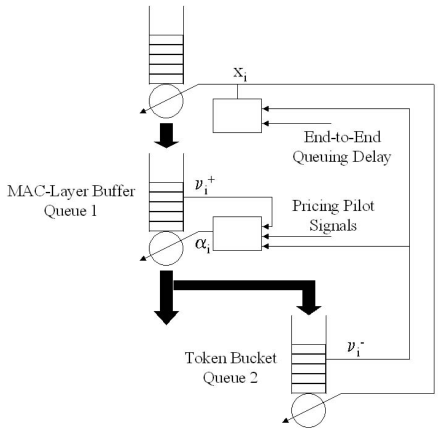

Fig. 5. MAC Layer Buffer and Token Bucket Structure

$\sum_{l=1}^{L} \frac{g_{il}}{g_{ib(i)}}\mu_l \backsimeq \frac{E_{TR}^{PPS}}{E_T^P(b(i))}$ where E_{TR}^{PPS} and $E_T^P(b(i))$ are quantities which can be measured locally by mobile i (see [18] and [21] for more details).

The practical scheme to compute the wired link signals $\underline{\lambda}$ and the corresponding aggregate signals $\underline{q}$ is well understood since Eqn (3) has a well-known interpretation in terms of queue delay at each link [5], [15], [16]. When dealing with a discrete-time system, however, the usual differential equation for queueing delay $\dot{\lambda}_j = \frac{1}{C_j}(\sum_{i=1}^{M} x_i\psi_{ij} - C_j)$ becomes $\Delta\lambda_j = \frac{\Delta t}{C_j}(\sum_{i=1}^{M} x_i\psi_{ij} - C_j)$, where Δt is the time between successive updates. In other words, if we take $\xi = \frac{\Delta t}{C_j}$, then λ_j is the queueing delay at link j, and q_i is user i's end-to-end queueing delay in the wired IP network. Regarding wireless users running the modular algorithm in a real system, however, it is α_i, and not x_i, which determines the queuing delay at the intermediate wired links. As such, we propose to approximate $\Delta\lambda_j \approx \frac{\Delta t}{C_j}(\sum_{i=1}^{N} \alpha_i\psi_{ij} + \sum_{i=N+1}^{M} x_i\psi_{ij} - C_j)$.

It is left to address the practical implementation of the cross-layer coordination signals ν_i+ and ν_i^- in the context of the modular algorithm. It is interesting to note that the the queueing delay interpretation above also applies here. When $\zeta_1 = \frac{\Delta t}{\alpha_i}$ and $\zeta_2 = \frac{\Delta t}{x_i}$, Eqns (5) and (6) are similar to the queueing delay equations if we think of two imaginary queues (see Figure 5): one with input traffic rate x_i and capacity α_i (Queue 1), and one with input traffic rate α_i and capacity x_i (Queue 2). Thus we choose $\zeta_1 = \frac{\Delta t}{\alpha_i}$ and $\zeta_2 = \frac{\Delta t}{x_i}$ to ensure that the quantities ν_i^+ and ν_i^- can be interpreted as the delays associated with Queues 1 and 2, respectively. This configuration is shown in Figure 5, and is similar to the concept of token buckets (see [6], [11], and references therein). Every time the transport layer sends traffic to Queue 1, it empties the same amount of traffic from Queue 2. Similarly, every time the MAC layer removes traffic for transmission from Queue 1, it adds the same amount of traffic to Queue 2. Queue 1 is now our actual link, and Queue 2 is our token bucket. The difference between this and the traditional token bucket is that we do not use the token bucket to regulate service rate, but instead use it to keep track of the mismatch between two rates, x_i and α_i.

Thus far, we have explained how the addition of a MAC-layer buffer and token bucket facilitates an online and practical computation of $\underline{\nu}^+$ and $\underline{\nu}^-$. In reality, the addition of a MAC-layer buffer and token bucket has a three-fold impact: 1) it eliminates the need for explicit computation of the cross-layer coordination signals, 2) it creates a natural distributed and adaptive priority scheme based on the MAC buffer and token bucket length (as $\nu_i = \nu_i^+ - \nu_i^-$ increases user i's local rule becomes more aggressive), and 3) it allows the transport layer protocol to unconsciously take interference levels into account *without any major modification of current protocols*. In order to understand the third impact, recall that TCP Vegas uses end-to-end queueing delay as its feedback mechanism [15]. This quantity is obtained by measuring the round-trip-time and subtracting the propagation delay. Looking at Eqn (7), we notice that the quantity $q_i + \nu_i$ is nothing more than the end-to-end queueing delay ($q_i + \nu_i^+$) minus the token bucket delay (ν_i^-). In other words, the only necessary modification to current transport layer protocols is to subtract the token bucket delay (in addition to the propagation delay) from the round-trip time!

B. Convergence Result

Typically, the convergence of gradient projection algorithms is dependent upon the step-size being "small enough" (see [3], pages 212-215). Recall, however, that our interpretation of the Lagrangian multipliers as queueing delays was based on choosing the step size as $\frac{\Delta t}{C}$, where C is the link capacity or service rate at a queue. As a result, to guarantee convergence we must simply run the algorithm "fast enough." This means that the convergence of both the one-shot and modular algorithms is dependent upon the time-scale of the distributed feedback loops shown in Figures 3 and 4.

With these issues in mind, we present the following theorems regarding the convergence of the one-shot and modular algorithms. The proof of these theorems can be found in [19].

Theorem 1: There exist values β_0 and δ_0 such that for all $\beta < \beta_0$ and for all $\delta < \delta_0$, if $\xi_j = \frac{\delta}{C_j}$ $\forall j$ then the one-shot distributed algorithm described by Eqns (2)-(4) converges to the solution to Problem P.

Theorem 2: There exist values β_0 and δ_0 such that for all $\beta < \beta_0$ and for all $\delta < \delta_0$, if $\xi_j = \frac{\delta}{C_j}$, $\zeta_1^{(i)} = \frac{\delta}{\alpha_i}$, and $\zeta_2^{(i)} = \frac{\delta}{x_i}$ then the modular algorithm described by Eqns (2)-(3) and Eqns (5)-(8) converges to the solution to Problem P'.

VI. Conclusions

In this paper, we have developed two sets of algorithms for optimal resource allocation in variable-rate CDMA networks. The first algorithm is a one-shot approach which merges the functionality of the transport and MAC layers, while the second is a modular approach which attempts to coordinate the two separate layers. We show that the addition of queues between the transport and MAC layers actually facilitates the coordination necessary for the modular approach with only minimal modification of current protocols. Finally, we discuss how the use of queueing delay as a congestion feedback mechanism impacts the convergence results of both algorithms.

The most important avenue of future research is a detailed comparison of the one-shot and modular algorithms. Although we know these algorithms will converge to the same cross-layer optimal rate assignment (i.e. the solution to Problem P), the dynamic behaviors of these algorithms will almost certainly differ. Early simulation results indicate that the modular algorithm tends to result in longer end-to-end queueing delays, due largely to the addition of the MAC layer queues for each wireless user. On the other hand, the structure of the modular algorithm allows the MAC and transport layer updates to be run at different time-scales. We believe that the choice of parameters (e.g. Δt, σ, etc) will play an important role in this trade-off in particular, and in the dynamic behavior of the algorithms in general.

Acknowledgments

This work was supported in party by the National Science Foundation CAREER award No. CNS-0347961.

References

[1] 3rd Generation Partnership Project 2 (3GPP2). CDMA2000 high rate packet data air interface specification. Technical Report CS20024, www.3gpp2.org, 2001.

[2] H. Balakrishnan, S. Seshan, E. Amir, and R.H. Katz. Improving TCP/IP performance over wireless networks. In *Proceedings of the First ACM Conference on Mobile Computing and Networking*, 1995.

[3] D.P. Bertsekas and J.N. Tsitsiklis. *Parallel and Distributed Computation*. Prentice-Hall Inc., 1989.

[4] J.F. Chamberland and V.V. Veeravalli. Decentralized dynamic power control for cellular CDMA systems. *IEEE Transactions on Wireless Communications*, 2(3):549–559, May 2003.

[5] M. Chiang. To layer or not to layer: Balancing transport and physical layers in wireless multihop networks. In *Proceedings of IEEE INFOCOM*, 2004.

[6] R.L. Cruz. A calculus for network delay I: Network elements in isolation. *IEEE Transactions on Information Theory*, 27(1):114–131, Jan 1991.

[7] R.L. Cruz and A.V. Santhanam. Optimal routing, link scheduling and power control in multi-hop wireless networks. In *Proceedings of IEEE INFOCOM*, 2003.

[8] T. ElBatt and A. Ephremides. Joint scheduling and power control for wireless ad hoc networks. *IEEE Transactions on Wireless Communications*, 3(1):74–85, January 2004.

[9] D. Julian, M. Chiang, D. O'Neill, and S. Boyd. QoS and fairness constrained convex optimization of resource allocation for wireless cellular and ad hoc networks. In *Proceedings of IEEE INFOCOM*, pages 477–486, 2004.

[10] F. Kelly. Mathematical modelling of the Internet. In B. Engquist and W. Schmid, editors, *Mathematics Unlimited - 2001 and Beyond*, pages 685–702. Springer-Verlaq, 2001.

[11] S.S. Kunniyur and R. Srikant. An adaptive virtual queue (AVQ) algorithm for active queue management. *IEEE/ACM Transactions on Networking*, 12(2):286–299, April 2004.

[12] R.J. La and V. Anantharam. Utility-based rate control in the Internet for elastic traffic. *IEEE Transactions on Networking*, 10(2):272–286, April 2002.

[13] R. Leelahakriengkrai and R. Agrawal. Scheduling in multimedia CDMA wireless networks. *IEEE Transactions on Vehicular Technology*, 52(1), Jan 2003.

[14] X. Lin and N.B. Shroff. The impact of imperfect scheduling on cross-layer rate control in wireless networks. In *To Appear in the Proceedings of IEEE INFOCOM*, 2005.

[15] S. H. Low and D. E. Lapsley. Optimization flow control, I: basic algorithm and convergence. *IEEE/ACM Transactions on Networking*, 7(6):861–874, Dec 1999.

[16] J. Mo and J. Walrand. Fair end-to-end window-based congestion control. *IEEE/ACM Transactions on Networking*, 8(5):555–567, Oct 2000.

[17] T. Nandagopal, T. Kim, X. Gao, and V. Bharghavan. Acheiving mac layer fairness in wireless packet networks. In *ACM Mobicom 2000*, 2000.

[18] J. Price and T. Javidi. Decentralized rate assignments in a multi-sector CDMA network. Technical report, University of Washington, http://www.ee.washington.edu/techsite/papers.

[19] J. Price and T. Javidi. Distributed rate assignments for simultaneous interference and congestion control in CDMA-based wireless networks. Technical report, http://fleece.ucsd.edu/ tjavidi/mypapers.html.

[20] J. Price and T. Javidi. Cross-layer (mac and transport) optimal rate assignmentin cdma-based wireless broadband networks. In *Asilomar Conference on Signals, Systems and Computers*, pages 1044–1048, 2004.

[21] J. Price and T. Javidi. Decentralized and fair rate control in a multi-sector CDMA system. In *Proceedings of the Wireless Communications and Networking Conference*, 2004.

[22] V.A. Siris and C. Courcoubetis. Resource control for loss-sensitive traffic in CDMA networks. In *Proceedings of INFOCOM*, pages 2790–2799, 2004.

[23] C. Touati, E. Altman, and J. Galtier. Fair power and transmission rate control in wireless networks. In *Global Telecommunications Conference*, pages 1229–1233, 2002.

[24] M. Xiao, N.B. Shroff, and E.K.P. Chong. A utility-based power-control scheme in wireless cellular systems. *IEEE/ACM Transactions on Networking*, 11(2):210–221, April 2003.

[25] H. Yaiche, R. R. Mazumdar, and C. Rosenberg. A game theoretic framework for bandwidth allocation and pricing in broadband networks. *IEEE/ACM Transactions on Networking*, 8(5):667–678, October 2000.

[26] R. D. Yates. A framework for uplink power control in cellular radio systems. *IEEE Journal on Selected Areas in Communications*, 13(7):1341–1347, Sept 1995.