Hierarchical SIR and Rate Control on the Forward Link for CDMA Data Users under Delay and Error Constraints

Lei Song, *Student Member, IEEE* and Narayan B. Mandayam, *Senior Member, IEEE*
Wireless Information Network Laboratory (WINLAB)
Department of Electrical & Computer Engineering
Rutgers University, Piscataway, NJ 08854

Abstract - In this paper, we study SIR and rate control for CDMA data users on the forward link under average or peak power constraints. The QoS for data users is specified by delay and error rate constraints as well as a family of utility functions representing the throughput and fairness among the data users. It is found that the optimal SIR and rate control algorithm has an hierarchical structure which can be easily implemented in a distributed manner. The SIR targets can be adjusted independently by the mobiles using information specific to the individual users. The data rates can be adjusted jointly by the base station based on limited feedback from the mobiles. We also propose a 2-level iteration algorithm for both the mobile and the base station to efficiently compute the SIR and data rates. Our results show that a flexible tradeoff between total system throughput (sum of rates achieved) and fairness (similarity in data rates) can be achieved by choosing appropriate utility functions used in this scheme.

Keywords: Wireless Data, CDMA, Rate and SIR Control, Hierarchical Architecture.

This work is supported in part by the NSF under Grant no. NCR 97-06036. This work was presented in part at ICC'2000, New Orleans, LA, June 2000.

1 Introduction

In the upcoming third generation (3G) CDMA systems such as [1] and future generations of wireless systems, it is expected that the increasing popularity of wireless data applications such as Internet access will make the traffic more asymmetric with the majority being carried on the forward link. This may result in a forward link limited system. As a consequence, efficient utilization of system resources on the forward link is becoming more important. System resources such as power, signal to interference ratio (SIR) and data rate need to be efficiently controlled to maintain overall system feasibility and optimize network performance measures such as delay and throughput for data.

Much work has been done on SIR and rate control for integrated voice/data CDMA systems. In [2], the ratio of power allocation for voice and data users is studied. In [3–5], different types of data services are characterized by different minimum SIR requirements and rate control algorithms are designed to provide admission control. [6] studies power and rate allocation based on channel states. Though most of the work in the literature has primarily considered the reverse link, the results can be extended to the forward link. Forward link specific SIR and rate control can be found in [7, 8] where some computation needs to be performed at the base station. A detailed description of the data rate adaption procedures used in practical systems such as in [1,9] can be found in [10]. Some of the earlier approaches are based on assumptions that the quality of service (QoS) for a data user can be characterized by a minimum SIR requirement. Some others assume that an errored frame will be retransmitted repetitively until it goes through.

In practical systems [1, 9], a major concern is reliable end-to-end transmission. Since the transmission often involves the PSTN (Public Switched Telephone Network), an end-to-end error recovery scheme such as TCP (the Transmission Control Protocol) is usually implemented in the mobile and the interworking function (IWF) at the CDMA system side [11]. However, TCP can not directly interact with the CDMA system physical layer [12]. TCP was originally designed for wireline application and its error recovery scheme only works well on channels with very low error rate. Since the CDMA physical layer does not distinguish voice and data, it provides a lossy channel to data users as well (usually having a frame error rate of 10^{-2}). If TCP is implemented directly on top of the physical layer, such a lossy channel will cause excessive TCP time-outs and retransmissions and results in very low data throughput. Since data is more tolerant to delay than voice, this characteristic can be exploited to facilitate the interaction between the radio physical layer and TCP. It is achieved by the insertion of a function called the Radio Link protocol (RLP) [11] between the physical layer and the TCP that trades a certain delay tolerance for a more reliable channel transmission. The RLP retransmits an errored frame for only a limited number of times. The number of allowable retransmissions is chosen such that the delay introduced by the retransmission is tolerable. The residual error rate after the retransmission, which we call the Post RLP Error Rate (PRER) can be made low enough (such as less than 10^{-5}) such that the TCP can perform reliable end-to-end data transmission with its own data recovery scheme.

In this paper, we study SIR and rate control for the forward link under constraints on average total transmit power (due to the power limited forward link) and the Post RLP Error Rate (due to the TCP requirement on end-to-end recovery). We characterize the QoS for a data user by a utility function that is a measure of the achieved throughput and also reflects the degree of fairness among the users. The proposed algorithm computes the SIR and data rate assignments that can maximize the sum of utilities for all data users under delay and PRER constraints. The sum of utilities in the system is a measure of total throughput that also takes into account the fairness constraints that prevent all resources from being allocated to users with better channels. We also extend the results to the case of a peak power constraint on the forward link.

The proposed algorithm has an hierarchical structure that can be implemented in a distributed fashion and requires very limited exchange of information between the base station and the mobile stations. The SIR target (on the forward link) is selected individually by each mobile based on its own channel condition to achieve a certain frame error performance. The data rates for all the data users (on the forward link) are decided jointly by the base station to maintain network stability and optimize the overall system performance. Further, for both the mobile and the base station, we propose a 2-level hierarchical algorithm that can effectively compute the optimal SIR and rate assignments with guaranteed convergence.

2 System Model

We consider a CDMA system where the forward link SIR can be controlled by the mobiles. Each mobile decides its own SIR target and then issues a sequence of power control bits to instruct the base station to adjust its transmit power to the mobile to meet the SIR target[1]. If the received SIR falls bellow the target, the mobile instructs the base station to increase its power; otherwise, it instructs the base station to reduce the power. In this paper, we assume that the underlying power control algorithm can achieve the SIR target[2]. The SIR target is determined by the mobile to achieve a certain frame error performance based on its forward link channel information. We assume that the channel information is given in the form of the forward link frame error rate (FER) for a SIR target, i.e., for a SIR target γ_i, the corresponding FER $f_i(\gamma_i)$ is known. We further assume that the function $f_i(\cdot)$ is analytic, monotonically decreasing and convex, which are consistent with observations in practice [15].

[1] In this paper, the signal to interference ratio (SIR) is SIR per bit which includes the processing gain. Note that though a mobile may require a very high SIR, the base station can assign the mobile with a lower rate, (i.e., a larger spreading gain) which only requires a lower transmit power. Thus the transmit power can still be under control by the base station even though the mobiles may choose their SIR targets.

[2] Power control algorithms such as the up/down power control used in practical systems [9] have been shown to be able to achieve the SIR target [13, 14].

2.1 Data Transmission Model Assumptions

We assume that all the data users are synchronous at the frame level and data is transmitted in segments as shown in Figure 1. The data rate for a user remains the same throughout a segment which lasts for multiple frames[3]. The RLP retransmits an errored frame up to M times such that the constraint on delay is met. We assume that once the base station knows a frame was in error, it retransmits the errored frame at the next available frame slot. For example, in IS-2000 [1], the length of a traffic frame is 20ms while the length of a control frame is 5ms and thus retransmission can be done with only one traffic frame delay. If an errored frame is scheduled to be retransmitted at the end of a data segment, it will then be transmitted as a new frame at the beginning of the next segment. For the original transmission of a frame from user i, the mobile sets a target SIR $\gamma_{i,1}$. If the frame is received in error, the mobile sends a NAK to the base station using the reverse link control channel asking for retransmission of this frame. For the jth retransmission of a frame, the SIR target is set at $\gamma_{i,j}$. The SIR vector $\underline{\gamma}_i = \{\gamma_{i,1}, ..., \gamma_{i,M}\}$ represents the SIR targets to be used in the data segment. The data rate R_i remains the same throughout the whole data segment. We further require that the segments are synchronized for all data users.

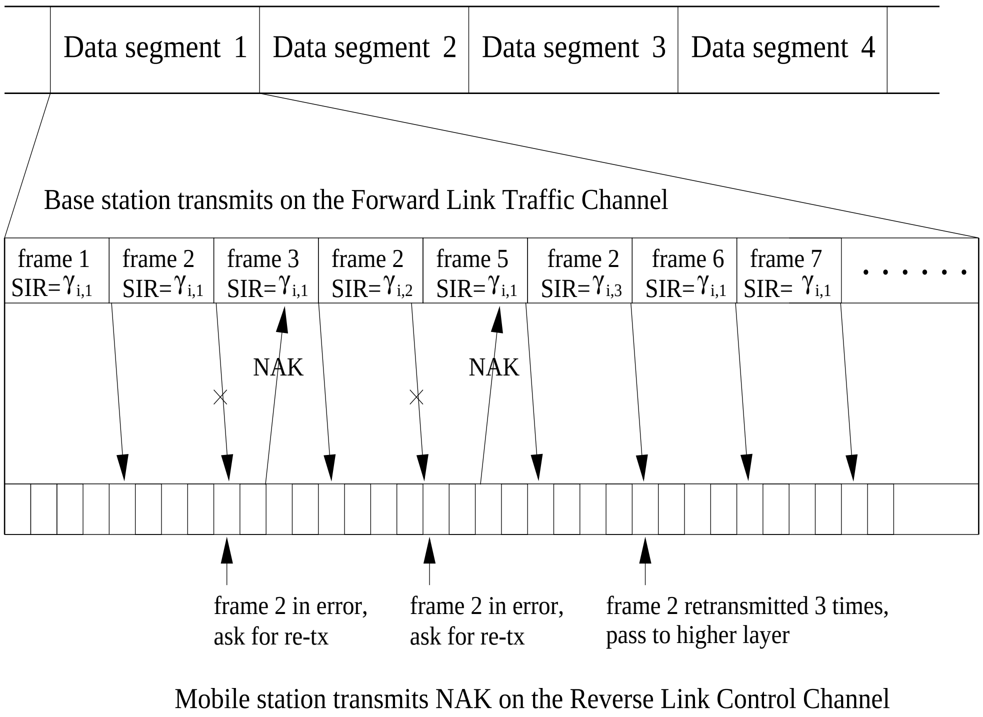

Figure 1: Example of the data transmission

[3]This is achievable in 3G systems since data can be transmitted on the supplemental channels and the signaling is done on the dedicated control channels [1]. No dim and burst scheme is needed for in-band signaling on the supplemental channel [1, 9].

2.2 QoS for Data Users

In this section, we define QoS for data users that we use in designing the SIR and rate control algorithm. It is well known that QoS for voice is well characterized by a target SIR. However, data users are characterized by performance requirements that are very different from voice. Such requirements may include system delay, error performance, data throughput and the degree of fairness among the users. We assume that the delay tolerance is specified by the limitation on the number of allowable retransmissions. However, the performance requirements are specified as a Post-RLP Error Rate (PRER) target that needs to be met. Within the constraints of delay and PRER, each individual data user prefers a higher throughput, i.e., achievable data rate. However, a certain degree of fairness (similarity in the data rates for different users) among the users is often desirable from a network point of view. Thus, given that the delay and PRER constraints are met, the QoS for data users should reflect the throughput and fairness.

There are many ways to define this QoS. Recently, there has been a lot of work on using utility functions to describe QoS for data in wireline [16, 17] and wireless [18–20] communications. In this paper, we adopt the utility function based approach. We define the utility function $U_i(R_i)$ for user i as a function of the data rate R_i. We assume that the utility function is concave and monotonically increasing. The concavity of $U_i(\cdot)$ imposes a higher penalty on a lower data rate. By adjusting the shape of $U_i(\cdot)$, a flexible tradeoff can be achieved between the overall optimality (total system data rate) and fairness (data rate discrepancy). This point will be illustrated when we present the numerical results in Section 5.

3 Problem Formulation

Consider a CDMA system with N data users. The control algorithm will find a set of SIR and rate parameters for each data user at the beginning of a data segment such that the total utility is optimized for the segment. The constraints are that (a) the average total forward link power for all the data users should be within a limit C; and (b) for each individual user, the PRER should be within a limit ϵ_0.

Mathematically, the optimization problem (denoted as problem (A)) can be formulated as choosing the vector $S_i^* = \{\underline{\gamma}_i^*, R_i^*\} = \{\gamma_{i,1}^*, ..., \gamma_{i,M}^*, R_i^*\}$ for $i = 1, ..., N$, such that

$$\max_{S_1,\ldots,S_N} \quad \sum_{i=1}^{N} U_i(R_i) \tag{A.1}$$

$$\text{subject to} \quad \sum_{i=1}^{N} \bar{p}_i \leq \ C \tag{A.2}$$

$$PRER_i \ \leq \ \epsilon_0 \ \text{ for } i = 1, ..., N \tag{A.3}$$

$$\gamma_{min} \leq \gamma_{i,m} \leq \gamma_{max} \text{ for } i = 1, ..., N, \ m = 1, ..., M, \tag{A.4}$$

$$R_{min} \leq R_i \leq R_{max} \ \text{ for } i = 1, ..., N, \tag{A.5}$$

Equation (A.2) is the constraint on the average power. $\bar{p}_i = E[p_i]$ is the average power for user i, where p_i is the base station transmit power for user i in a power control group(PCG). Note that a data segment contains multiple frames and each frame contains many PCGs. We assume that the transmit power p_i is controlled by a power control algorithm to achieve the SIR target for user i. Note that p_i is a random variable due to the variations in the channel that are modeled as a stochastic process. The power vector $\underline{\bar{p}} = [\bar{p}_1, ..., \bar{p}_N]$ can be related to the SIR and rate assignments $\{\underline{\gamma}_1, ..., \underline{\gamma}_N, \underline{R}\}$ as follows. Let X_i be the average SIR per bit used by user i in the data segment. X_i is related to $\underline{\bar{p}}$ as,

$$X_i = \frac{W}{R_i} \cdot \frac{E[h_i]E[p_i]}{E[h_i]\sum_{j=1,j\neq i}^{N} E[p_j] + I_i} \approx \frac{W}{R_i} \cdot \frac{E[h_i]E[p_i]}{E[h_i]\sum_{j=1}^{N} E[p_j] + I_i} = \frac{W}{R_i} \cdot \frac{\bar{p}_i}{\sum_{j=1}^{N} \bar{p}_j + I_i/\bar{h}_i} \quad (1)$$

where h_i is the link gain at every PCG (modeled as a random variable), $\bar{h}_i = E[h_i]$ is the average over the data segment and $\bar{p}_j = E[h_j]$. I_i is the other-cell interference plus thermal noise variance seen by user i. W is the spreading bandwidth. We have assumed in the above that on an average, the transmit powers and channel gains are statistically independent, i.e., $E[h_i p_i] = E[h_i]E[p_i]$. Note that the intracell interference is not ignored in the above formulation. This reflects the non-orthogonality of forward link users in a multipath environment. An approximation is made in (1) by dropping the restriction $j \neq i$. The approximation is valid when there are a large number of users in the system without a single dominant one [21]. In other words, the transmit power for any one user is much smaller than the total transmit power.

Based on equation (1), the power constraint (A.2) in the optimization problem can be rewritten as a constraint on X_i and R_i, for $i = 1, ..., N$. Treating the approximation in equation (1) as an equality, we have

$$\bar{p}_i = \frac{X_i R_i}{W}(\sum_{j=1}^{N} \bar{p}_j + I_i/\bar{h}_i). \quad (2)$$

Summing up equation (3) for all users,

$$\sum_{i=1}^{N} \bar{p}_i = \sum_{i=1}^{N} \frac{X_i R_i}{W} \sum_{j=1}^{N} \bar{p}_j + \sum_{i=1}^{N} \frac{X_i R_i I_i}{W \bar{h}_i}. \quad (3)$$

Thus, constraint (A.2) can be written as:

$$\sum_{i=1}^{N} \bar{p}_i = \frac{\sum_{i=1}^{N} \frac{X_i R_i I_i}{W \bar{h}_i}}{1 - \sum_{i=1}^{N} \frac{X_i R_i}{W}} \leq C. \quad (4)$$

Simplifying equation (4),

$$\sum_{i=1}^{N} \frac{I_i}{W \bar{h}_i} \cdot X_i R_i \leq C - \sum_{i=1}^{N} \frac{C}{W} \cdot X_i R_i \quad (5)$$

Then,

$$\sum_{i=1}^{N} \frac{I_i + C\bar{h}_i}{W\bar{h}_i} \cdot X_i R_i \leq C \tag{6}$$

Let $\alpha_i = \frac{I_i + C\bar{h}_i}{W\bar{h}_i}$. Thus we see that the power constraint (A.2) can be written equivalently as the following constraint on the average SIR per bit X_i and the data rate R_i,

$$\sum_{i=1}^{N} \alpha_i X_i R_i \leq C, \quad i = 1, ..., N \tag{7}$$

Since we assume that the underlying power control mechanism can achieve the SIR targets, the average achieved SIR for a given SIR target will be equal to the target[4]. Thus, X_i can be related to the SIR target assignments as follows:

$$X_i = \sum_{m=1}^{M} \gamma_{i,m} q_i(m), \tag{8}$$

where $q_i(m)$ is the probability that the frame to be transmitted from user i has been transmitted $m-1$ times. In other words, $q_i(1)$ is the probability of the transmission of a previously untransmitted frame. $q_i(M)$ is the probability the frame has been transmitted $m-1$ times. $q_i(m)$ can be found from the Markov chain shown in Figure 2 as:

$$q_i(m) = \frac{\prod_{j=1}^{m-1} f_i(\gamma_{i,j})}{1 + \sum_{j=1}^{M-1} \prod_{k=1}^{j} f_i(\gamma_{i,k})}. \tag{9}$$

Note that the Post-RLP error rate for user i is given as:

$$PRER_i = \prod_{m=1}^{M} f_i(\gamma_{i,m}). \tag{10}$$

Further, we assume that the system is feasible such that the constraints (A.2) and (1) are not mutually exclusive, i.e., there exists at least one SIR and rate assignment meeting both the power and PRER constraints[5]. Thus, as a necessary condition, if the users are given the maximum SIR target for the initial transmission and possible retransmissions, the achieved PRER should be within the limit ϵ_0. Furthermore, we also assume the problem is not trivial

[4]In practice, the power control algorithm controls the SIR around the SIR target. We assume that the speed of convergence of the power control algorithm is much faster (e.g., 1500 Hz adjustment speed in [22]) than the adaptation due to rate allocation (one adjustment over multiple frames). We also assume the residual power control error is small so that we can use the SIR target as the average SIR over a long period of time (this assumption has been shown to be valid even for the simple up/down power control algorithm [13, 14]).

[5]If there is no SIR and rate assignment that can meet the power and PRER requirements, the system is in an overload situation where more users are admitted than the system capacity. In this case, one or more users are usually dropped by the admission control algorithm to maintain the overall system performance [21].

where assigning the user with the minimum SIR γ_{min} for all transmissions does not achieve the PRER requirement, i.e,

$$\prod_{m=1}^{M} f_i(\gamma_{max}) < \epsilon_0 < \prod_{m=1}^{M} f_i(\gamma_{min}) \quad \text{for } i = 1, ..., N.$$

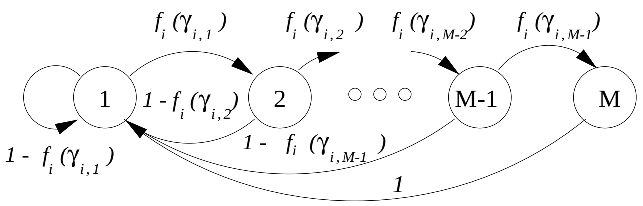

Figure 2: Markov chain for the retransmission of an errored frame

4 Hierarchical SIR and Rate Control

The problem (A) is a constrained optimization problem, which can be solved using standard search algorithms [23]. However, a direct search approach is not practical since it requires tremendous computational complexity to search through a very high dimensional space. Further, it also lacks a distributed structure that can be fit in the mobile station-base station architecture in CDMA systems. An effective approach to solve the above optimization problem is the hierarchical control method which often provides simple distributed solutions. First, we introduce two theorems (see the Appendix for proofs) to simplify optimization problem (A) into two sub-problems.

Theorem 1 *Let $S_i^* = \{\underline{\gamma}_i^*, R_i^*\}$ for $i = 1, ..., N$ be the solution of optimization problem (A). Then at least one of the following statements is true:*
(1): $R_i^* = R_{max}$ for all $i = 1, ..., N$, or,
(2): the total average power constraint in equation (A.2) is met with equality, i.e., $\sum_{i=1}^{N} \bar{p}_i = C$, and the PRER constraints in equation (1) are met with equality, i.e., $PRER_i = \epsilon_0$, for $i = 1, ..., N$.

Theorem 1 implies that the optimal SIR and rate vector S_i^* should result in the PRER being equal to ϵ_0 and a total average power of C except for the case where $R_i^* = R_{max}$ for all i. Intuitively, it means that if there is room in system power or PRER constraints, we should increase the rate R_i until $R_i = R_{max}$ or the constraints are met with equality. Since the underloaded case where $R_i^* = R_{max}$ for all $i = 1, ..., N$ is trivial, we will focus on the loaded case where all users can not be assigned with the maximum data rate.

Theorem 2 *Let $\underline{\gamma}_i^* = \{\gamma_{i,1}^*, ..., \gamma_{i,M}^*\}$ be the vector that minimizes the required X_i for user i (see equation (8)) , subject to the constraint that $PRER_i = \epsilon_0$. Suppose the system is not underloaded, i.e., there exists some user j, such that $R_j^* < R_{max}$. Then $S_i^* = \{\underline{\gamma}_i^*, R_i^*\}$ for $i = 1, ..., N$ is the solution of the optimization problem (A) for some $R_i^* \in [R_{min}, R_{max}]$.*

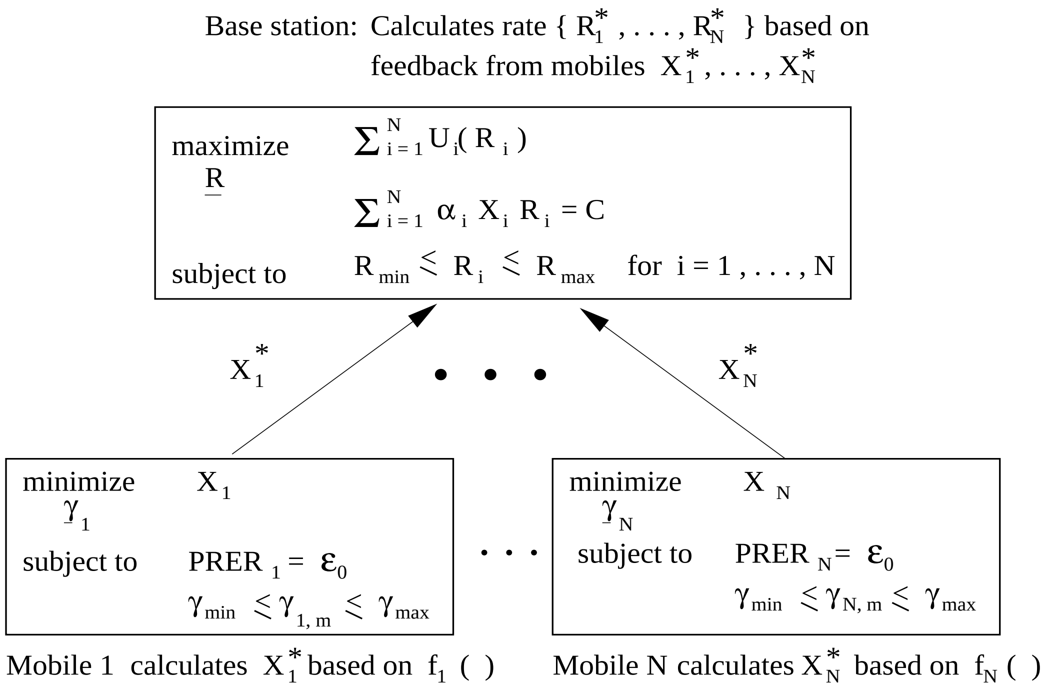

Figure 3: Hierarchical structure of the SIR and rate control algorithm

Theorem 2 states that the optimal SIR vector $\underline{\gamma}_i^*$ is the one that meets the PRER requirement and requires minimum average SIR per bit X_i. Intuitively, it means that to achieve the best total utility, each user's SIR should be assigned such that it requires the minimum average SIR to achieve the individual PRER constraint, i.e., $PRER_i = \epsilon_0$. Based on Theorem 1 and Theorem 2, the optimization problem (A) can be separated into two sub-problems with a hierarchical structure (see Figure 3). The first sub-problem (referred to as the mobile algorithm) follows directly from Theorem 2. The algorithm allows each mobile to determine the required downlink SIR target vector at which its individual signal should be received.

Problem (B) (Mobile Algorithm): For each user i, choose $\underline{\gamma}_i^* = \{\gamma_{i,1}^*, ..., \gamma_{i,M}^*\}$, such that

$$\min_{\underline{\gamma}_i} \quad X_i(\underline{\gamma}_i) \tag{B.1}$$

$$\text{subject to} \quad PRER_i = \epsilon_0 \tag{B.2}$$

$$\gamma_{min} \leq \gamma_{i,m} \leq \gamma_{max} \text{ for } m = 1, ..., M \tag{B.3}$$

We will show in Section 4.1 that Problem (B) can be solved in the mobile using only local information $f_i(\cdot)$ and hence it is referred to as the mobile algorithm.

Now consider Theorem 1, according to which the total average power should be set at the limit C, i.e., $\sum_{i=1}^{N} \bar{p}_i = C$ except for underloaded systems where the power is not limiting. Using this fact, we can now state the second sub-problem referred to as the base station algorithm as:

Problem (C) (Base Station Algorithm): Given the optimum SIR per bit $\underline{X}^* = \{X_1^*, ..., X_N^*\}$ from Problem 1, choose $\underline{R}^* = \{R_1^*, ..., R_N^*\}$, such that

$$\max_{\underline{R}} \quad \sum_{i=1}^{N} U_i(R_i) \tag{C.1}$$

$$\text{subject to} \quad \sum_{i=1}^{N} \alpha_i X_i^* R_i = C \tag{C.2}$$

$$R_{min} \leq R_i \leq R_{max} \quad for\ i = 1, ..., N \tag{C.3}$$

Note that in Problem (C), the SIR per bit X_i^* for all the mobiles are needed. Thus Problem (C) can only be solved jointly for all users at the base station, where the downlink rates $\underline{R}^*$ are chosen.

4.1 Mobile Algorithm

In this section, we present an efficient 2-level iteration algorithm to solve Problem (B) at each mobile. To derive a tractable solution, equation (8) is approximated as

$$X_i \approx \sum_{m=1}^{M} \gamma_{i,m} \prod_{j=1}^{m-1} f_i(\gamma_{i,j}) \tag{11}$$

The argument is that under normal operating conditions, $f_i(\gamma_i) \ll 1$. Thus the term $\sum_{j=1}^{N-1} \prod_{k=1}^{j} f_i(\gamma_{i,k})$ is much less than 1. Numerical results presented later in the paper will show that this approximation is reasonably accurate. The approximated optimization problem (denoted as problem (D)) is to choose $\underline{\gamma}_i^* = \{\gamma_{i,1}^*, ..., \gamma_{i,M}^*\}$ such that

$$\min_{\underline{\gamma}_i} \quad \sum_{m=1}^{M} \gamma_{i,m} \prod_{j=1}^{m-1} f_i(\gamma_{i,j}) \tag{D.1}$$

$$\text{subject to} \quad \prod_{m=1}^{M} f_i(\gamma_{i,m}) = \epsilon_0 \tag{D.2}$$

$$\gamma_{min} \leq \gamma_{i,m} \leq \gamma_{max} \quad \text{for} \quad m = 1, ..., M \tag{D.3}$$

Note that only local information $f_i(\ )$ corresponding to user i is needed in solving the problem. The problem (D) can be solved very efficiently in the mobile using an hierarchical control method as follows. The Lagrangian for the optimization problem (D) is:

$$L(\underline{\gamma}_i, \lambda) = L(\gamma_{i,1}, ..., \gamma_{i,M}, \lambda) = \sum_{m=1}^{M} \gamma_{i,m} \prod_{j=1}^{m-1} f_i(\gamma_{i,j}) + \lambda(\prod_{m=1}^{M} f_i(\gamma_{i,m}) - \epsilon_0).$$

We form a dual function as:

$$\phi(\lambda) = \min_{\underline{\gamma}_i} L(\underline{\gamma}_i, \lambda), \quad \gamma_{i,m} \in [\gamma_{min}, \gamma_{max}], \text{ for } m = 1, ..., M \tag{12}$$

The dual problem of the optimization problem (D) is to maximize the dual function $\phi(\lambda)$:

$$\max_{\lambda} \phi(\lambda) \tag{13}$$

Theorem 3 *There is no duality gap between optimizing the original problem (D) and maximizing the dual function (28).*

The proof of this theorem is given in the Appendix. Again, this dual problem can be solved using direct search which requires complex computation. However, using the classical goal-coordinate hierarchical control method [24], we can solve the dual problem efficiently using a 2-level iteration algorithm (see Figure 4). Thus the mobile algorithm can be implemented via a 2-level algorithm where one of the levels (level 1) chooses a set of targets $\underline{\gamma}_i$ for a given value of the Lagrange parameter λ and the other level (level 2) updates λ for a fixed set of target values in an effort to meet the PRER constraint. In the following, we briefly describe the two levels of the algorithm.

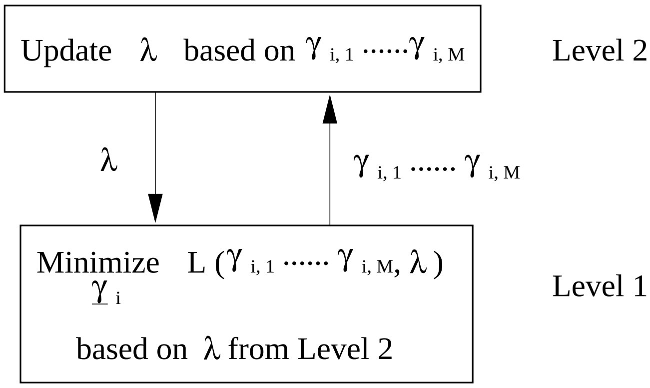

Figure 4: 2-level implementation of the mobile algorithm

Level 2 Algorithm: We update $\lambda(k+1)$ for a given $\underline{\gamma}_i(k)$ using a gradient search algorithm. We formulate the error at the iteration k as:

$$e(k) = \phi'(\lambda(k)) = \prod_{m=1}^{M} f_i(\gamma^*_{i,m}(k)) - \epsilon_0 \tag{14}$$

and then update $\lambda(k+1)$ as:

$$\lambda(k+1) = \lambda(k) + a\Delta(k)$$

where a is the step size and $\Delta(k)$ is the search direction at iteration k. The steepest descent method sets:

$$\Delta(k) = e(k) \tag{15}$$

The conjugate gradient method sets:

$$\Delta(k) = e(k) + \frac{e^2(k)}{e^2(k-1)}\Delta(k-1) \tag{16}$$

Level 1 Algorithm: We update $\underline{\gamma}_i(k)$ based on $\lambda(k)$ obtained from Level 2. The value of $\underline{\gamma}_i^*(k)$ is found by optimizing $L(\underline{\gamma}_i, \lambda(k))$ and can be given in closed-form as follows.

$$\gamma_{i,M}^* = \begin{cases} \gamma_{min} & \text{if } \; Y_i^M < f'(\gamma_{min}) \\ \gamma_{max} & \text{if } \; Y_i^M > f'(\gamma_{max}) \\ \frac{df_i}{d\gamma_i}^{-1}(Y_i^M) & \text{if } \; \text{otherwise,} \end{cases} \tag{17}$$

where

$$Y_i^M = -\frac{1}{\lambda(k)} \tag{18}$$

Note that (17) gives the optimum SIR value only for the Mth retransmission. The other SIR values $\gamma_{i,m}^*(k)$ can be obtained as:

$$\gamma_{i,m}^* = \begin{cases} \gamma_{min} & \text{if } \; Y_i(m) < f'(\gamma_{min}) \\ \gamma_{max} & \text{if } \; Y_i(m) > f'(\gamma_{max}) \\ \frac{df_i}{d\gamma_i}^{-1}(Y_i(m)) & \text{if } \; \text{otherwise,} \end{cases} \tag{19}$$

where

$$Y_i^m = -\frac{1}{\lambda(k)\prod_{j=m+1}^{M} f_i(\gamma_{i,j}^*(k)) + \sum_{j=m+1}^{M}\gamma_{i,j}^*(k)\prod_{l=j+1}^{M} f_i(\gamma_{i,l}^*(k))}, \;\; m = 1, ..., M-1. \tag{20}$$

The convergence property of the mobile algorithm is summarized in the following theorem. The proof is given in the Appendix.

Theorem 4 *The mobile algorithm (14) - (20) converges to a local optimum solution of the dual problem (12).*

Note that the basic idea of the mobile algorithm is to find the minimum average SIR required to achieve the PRER constraint. We can have a similar observation that A non-optimal SIR target which can achieve the PRER constraint will require a higher average SIR. As a result, the base station will consider it unfavorably and allocate a lower data rate. This point will be clear later in the numerical results section.

4.2 Base Station Algorithm

After the 2-level algorithm converges, the mobile will use the resulting SIR targets as the targets for the power control algorithm in the next data segment. The mobile calculates the minimum required average SIR per bit X_i^* from equation (8) and sends it to the base station for rate determination. The base station chooses the values of $\underline{R}^* = \{R_1^*, ..., R_N^*\}$ to optimize Problem (C) based on $\{X_1^*, ..., X_N^*\}$. The assumption here is that the SIR targets and rates are feasible and that the base station can adjust its transmit power and data rate to achieve the target for each mobile. Again, an hierarchical algorithm can efficiently calculate the solution (see Figure 5) with guaranteed convergence. The derivation is similar to the mobile algorithm in the last section. The algorithm is given as follows:

Level 2 algorithm: We update the Lagrange parameter $\mu(k+1)$ as:

$$\mu(k+1) = \mu(k) + a\Delta(k) \tag{21}$$

where a is the step size and $\Delta(k)$ is the searching direction calculated using the search error

$$e(k) = \phi'(\mu(k)) = \sum_{i=1}^{N} \alpha_i X_i^* R_i^*(k) - C$$

Level 1 algorithm: We update $R_i^*(k)$ based on the value $\mu(k)$ obtained from Level 2.

$$R_i^* = \begin{cases} R_{min} & \text{if } \; -\mu(k) > U'(\gamma_{min}) \\ R_{max} & \text{if } \; -\mu(k) < U'(\gamma_{max}) \\ \frac{dU}{dR_i}^{-1}(-\mu(k)) & \text{if otherwise} \end{cases} \tag{22}$$

The convergence property of the base station algorithm is summarized in the following theorem. The proof is given in the Appendix.

Theorem 5 *The base station algorithm (21) - (22) converges to the unique solution of the optimization problem (D).*

5 Numerical Results

In this section, we present some numerical results on the hierarchical SIR and rate control algorithm for wireless data users. For the purpose of illustration, we model the radio channel for different users by choosing the FER function $f_i(\cdot)$ as:

$$f_i(\gamma_i) = 1 - e^{-\frac{1}{F_i}}, \tag{23}$$

where F_i is the parameter representing the fade margin. In [25], it is shown that the FER for various channels can be approximated by changing the value of F_i. To reflect the difference between the channels, the fade margin can be further written as

$$F_i = \beta_i F_0 \gamma_i$$

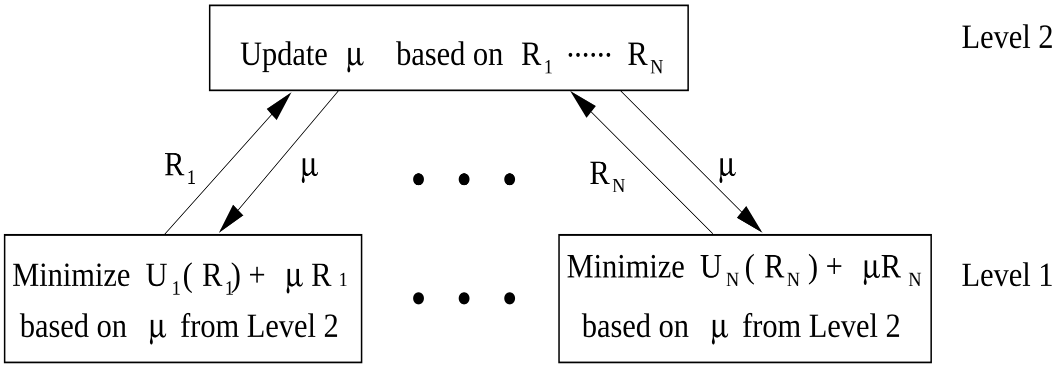

Figure 5: 2-level implementation of the base station algorithm

where F_0 is a constant and β_i is the parameter reflecting the power control efficiency [26]. For perfect power control, $\beta_i = 1$. While the results presented here are for the specific choices of $f_i(\cdot)$, the resource allocation approach works for any FER function that is analytic, convex and monotonically decreasing. In the simulation, we assume there are 6 users with $\beta_i = 0.2, 0.3, ...0.7$. $F_0 = 7.5$. The lower limit of SIR $\gamma_{min} = 3dB$ and upper limit of SIR $\gamma_{max} = 19dB$. For the forward link, the total average power limit C is set at 8 Watts. The link gains are $[-35\ -32\ -29\ -26\ -23\ -20]$ in dB and the interference levels are $[-13\ -16\ -19\ -22\ -25\ -28]$ in dB. Note that the link gains and corresponding interference levels are chosen based on a simple distance-based geometry for the purpose of illustration. The resource allocation approach presented here is more generally applicable to any cellular setting. The particular illustration presented here assumes a geographic distribution where user 6 is the closest to the base station and user 1 is the farthest away. From user 1 to user 6, link gains increase in 3 dB increments reflecting change in location (or equivalently change in channel quality). We also assume that from user 1 to user 6, the interference levels decrease since their locations are such that they are closer to the base station and further away from neighboring cells that create the interference. If there is no other-cell interference but only multipath, then I only represents the thermal noise, which could be the same for all users. Based on these parameters, we can see that user 6 has the best channel and user 1 has the worst channel.

The spreading bandwidth is 1MHz. The minimum rate for a data user R_{min} is 20K bits/sec and the maximum R_{max} is 200K bits/sec. The RLP retransmits an errored frame at most 3 times (i.e., $M = 3$) before it being passed to the higher layer for end-to-end error control. To ensure proper operation of the higher layer protocols, the Post-RLP error rate is kept less than 5×10^{-6}. The utility function we suggest is as follows:

$$U_i(R_i) = \frac{R_i}{R_{max}} - (\frac{R_i}{R_{max}})^{\rho} \tag{24}$$

The concavity can be guaranteed by choosing $\rho \leq 0$. The parameter ρ in the utility

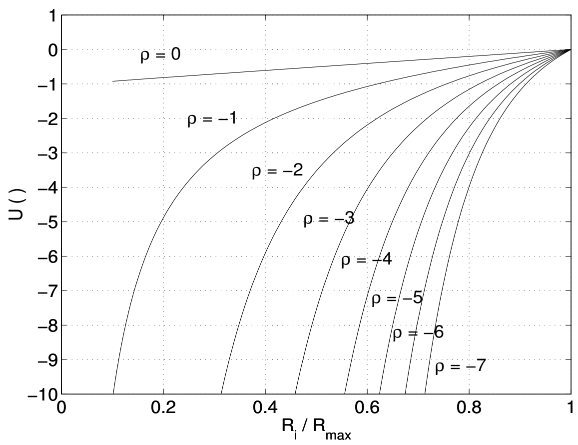

Figure 6: Utility function $U_i(\cdot)$ for different values of ρ

function $U_i(\cdot)$ can be used to balance the overall system performance such that the individual rate assignments for users are more equitable. The results for the problem using the approximate formulation in (12) is given in Table 1. The exact results are given in Table 2 . We can see that the approximation produces little degradation. However, if the condition $f_i(\gamma_i) \ll 1$ is not met, the approximation may cause a certain degradation. For example, if we increase the FER by allowing a higher PRER target $\epsilon_0 = 5 \times 10^{-2}$ and allowing a lower $\gamma_{min} = -7dB$, then, SIR per bit X_1 calculated using the approximation in (12) is 0.63 compared to the true optimum of 0.56. In this case, the average FER is 42.5%, far outside the range for normal operation. Thus, we believe, the approximation normally does not result in significant degradation. From the results, we can see that to maintain the PRER limit with lower average SIR per bit, a higher SIR is used for retransmitted frames. This may increase the first time FER. However, a higher γ_{min} can reduce this. The results also show that users with better channels require lower SIR targets to meet the PRER constraints.

The resulting rate assignments from the base station using different values of ρ is given in Table 3, where we can see that a lower value of ρ results in a fairer system but with lower total system performance. To address the tradeoff better, we also plot the sum and the variance of the normalized data rate $\tilde{R}_i = \frac{R_i}{R_{max}}$ in Figure 7. Note that if the total data rate is preferred, a higher value of ρ should be chosen and this will result in a high discrepancy of the data rates among the users. In the extreme case where $\rho = 0$, the total utilities $\sum U_i(R_i)$ becomes simply proportional to the sum of the data rates $\sum R_i$. In this case, power will be allocated to the users with better channels in a water-filling fashion. In other words, some users with better channels will be given the maximum data rates while the other users will be given the minimum. This is a scenario where the network operator only wants to maximize the total network throughput and does not care about the fairness among a class of users. Thus, we can see that a tradeoff between the total data rate and fairness can be achieved by adjusting the value of ρ.

Table 1: SIR assignments by the mobile

	user 1	user 2	user 3	user 4	user 5	user 6
$\gamma_{i,1}^*$ (in dB)	9.8	7.4	5.7	4.5	3.4	3.0
$\gamma_{i,2}^*$ (in dB)	18.6	15.7	13.6	12.0	10.8	9.0
$\gamma_{i,3}^*$ (in dB)	19	19	19	19	19	19

Table 2: SIR assignment from Problem 1

	user 1	user 2	user 3	user 4	user 5	user 6
$\gamma_{i,1}^*$ (in dB)	9.7	7.4	5.6	4.4	3.2	3.0
$\gamma_{i,2}^*$ (in dB)	18.8	15.8	13.8	12.2	10.8	8.9
$\gamma_{i,3}^*$ (in dB)	19	19	19	19	19	19

Table 3: Rate assignment by the base station (in Kbps)

ρ	R_1^*	R_2^*	R_3^*	R_4^*	R_5^*	R_6^*
0	20.0	25.8	200	200	200	200
-1	25.4	48.3	87.9	175.6	200	200
-2	31.1	47.6	68.4	96.3	135.2	195.4
-3	34.1	46.5	60.9	78.1	98.8	123.5
-5	36.4	44.8	53.0	63.2	73.8	85.3
-10	38.4	42.9	47.4	51.8	56.4	61.0
-15	39.0	42.2	45.1	48.0	50.9	53.7

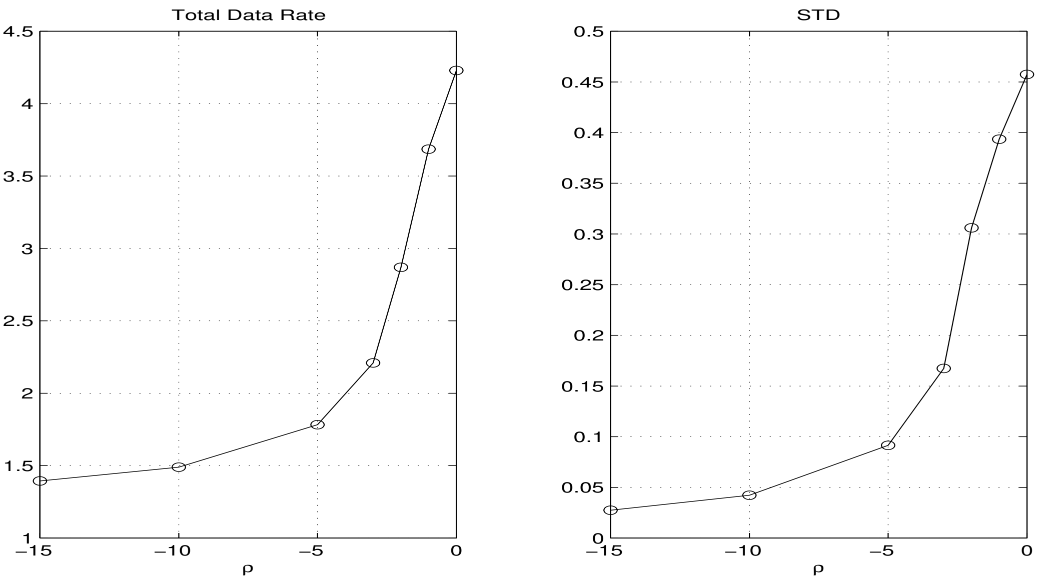

Figure 7: Normalized total data rate and variance for different values of ρ for the forward link

6 Extension To Peak Power Constraint

In a practical system, sometimes the maximum or peak power is more important than the average power for the forward link. The reason is that a very high short-term transmit power, even though it may not affect the average, may be very detrimental to the amplifier. Thus, it is sometimes useful to consider a constraint on the maximum power rather than on the average power. For an amplifier, the maximum power requirement is always specified as a short-term average power where the power measurement is obtained by averaging over a short time window. For example, the window of averaging is usually on the order of 10 10 ms or 20ms. Thus it is reasonable for us to consider the maximum power per frame here. The hierarchical approach we proposed here can be easily modified to solve this problem.

First, we notice that if the maximum power is considered, then the retransmission of a frame should use the same power as the original transmission. The reason is that if the powers are different, for example, $p_i(1) > p_i(2)$, we can increase $p_i(2)$ and make $p_i(2) = p_i(1)$. In this case, the peak power constraint is still met since it only counts the largest one of the set of power assignments used for original transmission and retransmission and the value does not increase by increasing $p_i(2)$ to $p_i(1)$. On the other hand, this lowers the PRER. Due to the continuity of all the functions involved, we can instead lower the maximum power and still achieve the PRER target. Thus, we can argue that for a user, it should use the same power for the original transmission and retransmission of a frame. Consequently, the only difference in this case is that the SIR targets remain fixed for the original transmission of a frame and its retransmissions.

The overall algorithm for the fixed SIR case has a similar hierarchical structure. The mobiles will still decide the SIR per bit X_i required for the PRER constraint. Let γ_i be the SIR used by user i. Its PRER is simply $\prod_{i=1}^{M} f_i(\gamma_i)$. Thus X_i can be found as

$$X_i = \gamma_i = f_i^{-1}(\epsilon_0^{1/M})$$

Once X_i is found for every user, it can be passed to base station where a rate determination algorithm similar to the earlier case can calculate the optimal rate assignment.

6.1 Numerical Results

In this section, we present numerical results for SIR and rate control under a maximum power constraint. The system considered in this example is identical to the one considered in Section 5 with the difference being that the maximum power is set to 8W instead of the average power. First, given the PRER $\epsilon_0 = 5 \times 10^{-6}$, the SIR per bit requirement for the six users is given in Table 4. Then, the optimal rates are calculated using the hierarchical rate assignment algorithm. For different values of ρ, the results are given in Table 5. We observed that the achieved data rates are reduced because now we have a tighter power constraint. In a practical system, if an amplifier is specified as a 8W (average power) amplifier, we may want to increase the peak power constraint to more than 8W. In this case, we also observe that achieving a better overall performance is usually at the cost of fairness, and vice-versa.

Table 4: SIR assignment by the mobile

	user 1	user 2	user 3	user 4	user 5	user 6
X_i	38.66	25.67	19.32	15.46	12.88	11.40

Table 5: Rate assignment by the base station

ρ	R_1 (kb/s)	R_2 (kb/s)	R_3 (kb/s)	R_4 (kb/s)	R_5 (kb/s)	R_6(kb/s)
0	20	20	20	20	20	87
-1	20	23.482	27.170	30.4520	33.4320	36.1940
-2	21.532	24.644	27.122	29.222	31.058	32.694
-3	22.616	25.042	26.906	28.446	29.768	30.936
-5	23.734	25.382	26.650	27.660	28.512	29.248

7 Conclusion

In this paper, we proposed an hierarchical SIR and rate control algorithm for CDMA data users on the forward link under constraints on delay, error rate and average or peak power. The QoS for data users is mapped onto a family of utility functions that reflect the tradeoff between total system throughput and fairness. The proposed algorithm uses an hierarchical control structure that is simple to implement in a distributed manner. Specifically, the data rates for the users are adjusted by the base station while the SIR targets are adjusted by the mobiles. The mobiles make SIR adjustments only based on information specific to each user. The base station makes the rate assignments jointly based on the feedback from the mobiles. Further, we proposed a 2-level iteration algorithm for both the mobile and the base station that can effectively calculate the SIR and data rates. The algorithm requires very limited information exchange and thus it is a very attractive candidate for resource control in high speed CDMA data systems. The approach presented here can also be generalized to include both peak and average power constraints simultaneously.

A Appendix

A.1 Proof of Theorem 1

Let the SIR and rate assignment $\{\underline{\gamma}_j^*, R_j^*\} = \{\gamma_{j,1}^*, ..., \gamma_{j,M}^*, R_j^*\}$, for $j = 1, ..., N$, be the solution of optimization problem (A). Then, $\underline{R}^* = \{R_1^*, ..., R_N^*\}$ is the conditional optimal rate assignment given the SIR assignment $\underline{\gamma}^* = \{\underline{\gamma}_1^*, ..., \underline{\gamma}_N^*\}$. In other words, $\underline{R}^*$ is the solution of the following optimization problem:

$$\max_{\underline{R}} \quad \sum_{i=1}^{N} U_i(R_i) \tag{E.1}$$

$$\text{subject to} \quad \sum_{i=1}^{N} \alpha_i X_i^* R_i \leq C \tag{E.2}$$

$$R_{min} \leq R_i \leq R_{max} \quad for\ i = 1, ..., N, \tag{E.3}$$

where X_i^* is the average SIR per bit calculated from equation (8) using the SIR assignment $\underline{\gamma}_j^*$. Suppose statement 1 in Theorem 1 is not true, i.e., $R_i^* < R_{max}$ for some user i and the power constraint (in statement 2 of Theorem 1) is not met with equality, i.e., $\sum_{j=1}^{N} \alpha X_j^* R_j^* < C$. We will now show that this results in a contradiction. Since $\sum_{j=1}^{N} \alpha X_j^* R_j^* < C$, the solution $\underline{R}^*$ must lie in the interior of the hyperplane formed by $\sum_{j=1}^{N} \alpha X_j^* R_j^* = C$. Therefore,

$$\frac{dU_i(R_i)}{dR_i}|_{R_i = R_i^*} = 0.$$

This contradicts the assumption that the function $U_i(\cdot)$ is monotonically increasing since that requires $\frac{dU_i(R_i)}{dR_i} > 0$ for all R_i. We can conclude that if $\{\underline{\gamma}_j^*, R_j^*\}$ is the solution of optimization problem (A), then either $R_i^* = R_{max}$ for all users or the power constraint is met with equality.

We will now proceed to show that either statement 1 is true or the PRER constraint of statement 2 in Theorem 1 is true. Let $\{\underline{\gamma}_j^*, R_j^*\}$ be the solution of (A). Suppose neither of the above is true, i.e., there exists a user l such that $R_l^* < R_{max}$ and there exists a user i such that $PRER_i^* = \prod_{m=1}^{M} f_i(\gamma_{i,m}^*) < \epsilon_0$.

First, consider the case where $\gamma_{i,M}^* > \gamma_{min}$. Since the function $f_i(\cdot)$ is continuous and monotonically decreasing, there exists $\tilde{\gamma}_{i,M} < \gamma_{i,M}^*$ such that

$$f_i(\tilde{\gamma}_{i,M}) \prod_{m=1}^{M-1} f_i(\tilde{\gamma}_{i,m}) < \epsilon_0,$$

where $\tilde{\gamma}_{i,m} = \gamma_{i,m}^*$ for all $m < M$. In other words, $\underline{\tilde{\gamma}}_i = \{\tilde{\gamma}_1, ..., \tilde{\gamma}_{M-1}, \tilde{\gamma}_M\} = \{\gamma_i^*, ..., \gamma_{M-1}^*, \tilde{\gamma}_M\}$. Using the SIR assignment $\underline{\tilde{\gamma}}_i = \{\tilde{\gamma}_1, ..., \tilde{\gamma}_M\}$, the required average SIR per bit $\tilde{X}_i$ for user i is given from equation (8) as

$$\begin{aligned}
\tilde{X}_i &= \frac{\tilde{\gamma}_{i,M} \prod_{j=1}^{M-1} f_i(\gamma_{i,j}^*) + \sum_{m=1}^{M-1} \gamma_{i,m}^* \prod_{j=1}^{m-1} f_i(\gamma_{i,j}^*)}{1 + \sum_{j=1}^{M-1} \prod_{k=1}^{j} f_i(\gamma_{i,k}^*)} \qquad (25)\\
&< \frac{\gamma_{i,M}^* \prod_{j=1}^{M-1} f_i(\gamma_{i,j}^*) + \sum_{m=1}^{M-1} \gamma_{i,m}^* \prod_{j=1}^{m-1} f_i(\gamma_{i,j}^*)}{1 + \sum_{j=1}^{M-1} \prod_{k=1}^{j} f_i(\gamma_{i,k}^*)} = X_i^*.
\end{aligned}$$

For user j, $j \neq i$, we always have $\tilde{X}_j = X_j^*$, since $\tilde{\gamma}_j = \gamma_j^*$. Thus,

$$\sum_{j=1}^{N} \alpha_j \tilde{X}_j R_j^* < \sum_{j=1}^{N} \alpha_j X_j^* R_j^* \leq C.$$

Thus, replacing $\underline{\gamma}^*$ with $\underline{\tilde{\gamma}}$ results in a lower average total power. As a result, $\{\underline{\tilde{\gamma}}, \underline{R}^*\}$ can not be the solution for problem (A). Since the assignment $\{\underline{\gamma}^*, \underline{R}^*\}$ achieves the same utility as the assignment $\{\underline{\tilde{\gamma}}, \underline{R}^*\}$, it is obvious that $\{\underline{\gamma}^*, \underline{R}^*\}$ is not a solution for problem (A), which contradicts our assumption.

Now consider $\gamma_{i,M}^* = \gamma_{min}$. Let m' be such that $\gamma_{i,m'}^* > \gamma_{min}$ and $\gamma_{i,k}^* = \gamma_{min}$ for any $k > m'$. Then, there exist $\tilde{\gamma}_{i,m'} < \gamma_{i,m'}^*$ such that $\prod_{k=1}^{M} f_i(\tilde{\gamma}_{i,k}) < \epsilon_0$, where $\tilde{\gamma}_{i,k} = \gamma_{i,k}^*$ for $k \neq m'$. Then, consider the probability that a new frame is being transmitted at any given time using the two SIR target sets. They must satisfy the following inequality:

$$\tilde{q}_i(1) = \frac{1}{1 + \sum_{j=1}^{M-1} \prod_{k=1}^{j} f_i(\tilde{\gamma}_{i,k})} < \frac{1}{1 + \sum_{j=1}^{M-1} \prod_{k=1}^{j} f_i(\gamma_{i,k}^*)} = q_i^*(1). \qquad (26)$$

where $\tilde{q}_i(1)$ and $q_i^*(1)$ are the probabilities of transmitting a previously untransmitted frame using SIR target sets $\underline{\tilde{\gamma}}_i$ and $\underline{\gamma}_i^*$, respectively. The above inequality states that if the SIR target for one retransmission is reduced while other targets remain the same, at any given time, a user is less likely to be transmitting a new frame.

According to equation (8),

$$
\begin{aligned}
\tilde{X}_i - X_i^* &= [\ (\tilde{q}_i(1) - q_i^*(1))(\gamma_{i,1}^* + \sum_{j=1}^{m-2} \gamma_{i,k}^* \prod_{k=1}^{j} f_i(\gamma_{i,k}^*)\ ] \\
&\quad + [\ \tilde{\gamma}_{i,m} \prod_{k=1}^{m-1} f_i(\gamma_{i,k}^*) \tilde{q}_i(1) - \gamma_{i,m}^* \prod_{k=1}^{m-1} f_i(\gamma_{i,k}^*) q_i^*(1)\ ] \qquad (27) \\
&\quad + [\ \gamma_{min}(\tilde{q}_i(m+1) - q_i^*(m+1))(1 + \sum_{j=m+1}^{M-1} \prod_{k=m+1}^{j} f_i(\gamma_{i,k}^*))\ ] \\
&< [\ \gamma_{min}(\tilde{q}_i(1) - q_i^*(1))(1 + \sum_{j=1}^{m-2} \prod_{k=1}^{j} f_i(\gamma_{i,k}^*))\ ] \\
&\quad + [\ \gamma_{min}(\tilde{q}_i(1) - q_i^*(1)) \prod_{k=1}^{m-1} f_i(\gamma_{i,k}^*)\ ] \\
&\quad + [\ \gamma_{min}(\tilde{q}_i(m+1) - q_i^*(m+1))(1 + \sum_{j=m+1}^{M-1} \prod_{k=m+1}^{j} f_i(\gamma_{i,k}^*))\ ] \\
&= \gamma_{min}\ [\ \tilde{q}_i(1)(1 + \sum_{j=1}^{m-2} \prod_{k=1}^{j} f_i(\gamma_{i,k}^*) + \prod_{k=1}^{m-1} f_i(\gamma_{i,k}^*)) \\
&\quad + \tilde{q}_i(m+1)(1 + \sum_{j=m+1}^{M-1} \prod_{k=m+1}^{j} f_i(\gamma_{i,k}^*))\ ] \\
&\quad - \gamma_{min}\ [q_i^*(1)(1 + \sum_{j=1}^{m-2} \prod_{k=1}^{j} f_i(\gamma_{i,k}^*) + \prod_{k=1}^{m-1} f_i(\gamma_{i,k}^*)) \\
&\quad + q_i^*(m+1)(1 + \sum_{j=m+1}^{M-1} \prod_{k=m+1}^{j} f_i(\gamma_{i,k}^*))\ ] \\
&= \gamma_{min}(1-1) = 0
\end{aligned}
$$

Thus we show that $\tilde{X}_i < X_i^*$ regardless of the SIR targets used in the last few retransmissions. Thus, it follows that

$$\sum_{j=1}^{N} \alpha_j \tilde{X}_j R_j^* < \sum_{j=1}^{N} \alpha_j X_j^* R_j^* \leq C.$$

Using the arguments similar to the above, we can conclude that $\{\underline{\gamma}^*, \underline{R}^*\}$ is not a solution for problem (A), and this contradicts our assumption. Hence, Theorem 1 must be true.

A.2 Proof of Theorem 2

Let $\{\underline{\gamma}_i^*, R_i^*\}$, for $i = 1, ..., N$, be the solution of problem (A). Suppose there exists a user i with $R_i^* < R_{max}$ such that $\tilde{\gamma}_{i,m} \neq \gamma_{i,m}^*$ for some m, where $\underline{\tilde{\gamma}}_i = \{\tilde{\gamma}_1, ..., \tilde{\gamma}_M\}$ minimizes the average SIR per bit under the constraint that $PRER_i = \prod_{m=1}^{M} f_i(\tilde{\gamma}_{i,m}) = \epsilon_0$. Then, from the assumption that $\underline{\tilde{\gamma}}_i$ minimizes SIR per bit, its corresponding $\tilde{X}_i$ must be smaller than the average SIR per bit X_i^* achieved using $\underline{\gamma}_i^*$. Using similar argument as in the proof of Theorem 1 above, we can show that the utility achieved by $\{\underline{\gamma}_i^*, R_i^*\}$, for $i = 1, ..., N$, is not the highest. This contradicts the assumption that $\{\underline{\gamma}_i^*, R_i^*\}$, for $i = 1, ..., N$, is the solution of problem (A). We can conclude that any solution of problem (A) must have $\underline{\gamma}_i^*$ that minimizes X_i under the PRER constraint for all users.

A.3 Proof of Theorem 3

Since we assume the system is feasible such that the optimization problem (A) is solvable, we have as a necessary condition the inequality $\prod_{m=1}^{M} f_i(\gamma_{max}) < \epsilon$. Thus, the *Slater condition* (see Corollary 28.2.1 in [27]), which is a sufficient condition for the existence of the Lagrange multiplier, is met. And this implies that no duality gap exists between the optimization problem (12) and its dual function (12) [23].

A.4 Proof of Theorem 4

The convergence of the algorithm can be shown as follows. Since $f_i(\cdot)$ is monotonically decreasing, according to the Level 1 algorithm (17), for a given λ, $\gamma_{i,M}^*$ is the only optimal solution. And consequently, $\underline{\gamma}_i^*$ is the only optimal solution given a particular λ. In other words, the function $\phi(\lambda)$ is well defined (i.e., a one to one mapping). Thus the Level 2 algorithm (14) performs a numerical search for the maximum of the one dimensional function $\phi(\lambda)$. Since the function $f_i(\cdot)$ is analytic, the function $\phi(\lambda)$ is smooth. Thus the search algorithm on Level 2 achieves a local optimum.

A.5 Proof of Theorem 5

Similar to the Theorem 3, there is no duality gap between the optimization problem (12) and maximizing the dual function defined as:

$$\phi(\mu) = \min_{\underline{R}} L(\underline{R}, \mu) = \min_{\underline{R}} \sum_{i=1}^{N} U(R_i) + \mu(\sum_{i=1}^{N} \alpha_i X_i R_i - C) \tag{28}$$

where $\underline{R} \in [R_{min},\ R_{max}]^N$. We will show that the function $\phi(\mu)$ is concave. Let μ_1 and μ_2 be two Lagrange parameters. Let ξ be a real number. Let $\underline{R}^*$ be the rate assignments that maximize $L(\underline{R}, \xi\mu_1 + (1-\xi)\mu_2)$. Then,

$$\begin{aligned}
\phi(\xi\mu_1 + (1-\xi)\mu_2) &= \sum_{i=1}^{N} U(R_i^*) + (\xi\mu_1 + (1-\xi)\mu_2)(\sum_{i=1}^{N} X_i R_i^* - CW) \qquad (29)\\
&= \sum_{i=1}^{N} U(R_i^*) + \xi\mu_1(\sum_{i=1}^{N} X_i R_i^* - CW) + (1-\xi)\mu_2(\sum_{i=1}^{N} X_i R_i^* - CW)\\
&\geq \xi[\sum_{i=1}^{N} U(R_i^*) + \mu_1(\sum_{i=1}^{N} X_i R_i^* - CW)]\\
&\quad +(1-\xi)[\sum_{i=1}^{N} U(R_i^*) + \mu_1(\sum_{i=1}^{N} X_i R_i^* - CW)]\\
&\geq \xi\phi(\mu_1) + (1-\xi)\phi(\mu_2)
\end{aligned}$$

Thus, the function $\phi(\mu)$ is concave. And consequently, the search algorithm (21) in Level 2 converges to the global optimum.

References

[1] TIA/EIA/IS-2000.2, *Physical Layer Standard for cdma2000 Spread Spectrum Systems*, Telecommunication Industry Association, 1999.

[2] Q. Shen and W. Krzymien, "Optimal Power Assignment for Maximum Throughput in CDMA Personal Communication Systems with Integrated Voice/Data Traffic," *Wireless Personal Communications*, vol. 8, pp. 277–289, Apr. 1998.

[3] L. Yun and D. Messerschmitt, "Variable Quality of Service in CDMA Systems by Statistical Power Control," in *IEEE International Conference on Communications*, 1995, vol. 2, pp. 713–719.

[4] N. Guo and S. Morgera, "The Grade of Service for Integrated Voice/Data Wireless DS-CDMA Networks," in *IEEE International Conference on Communications*, 1994, vol. 2, pp. 1104–1110.

[5] D. Jeong and W. Jeon, "Rate Controlled Data Transmission For IS-95 CDMA Networks," in *VTC'97*, May 1997, pp. 1567 – 1571.

[6] S. Oh and K. Wasserman, "Dynamic spreading gain control in multi-service cdma networks," *IEEE Journal on Selected Areas in Communications*, vol. 17, pp. 918 – 927, 1999.

[7] J. Rezaiifar, A. Makowski, and S. Kumar, "Optimal Control of Handoffs in Cellular Networks," in *VTC'95*, May 1995, vol. 1, pp. 887–891.

[8] J. Razavilar, R. Liu, and S. Marcus, "Optimal Rate Control in Wireless Networks with Fading Channels," in *VTC'99*, May 1999, vol. 1, pp. 807–813.

[9] TIA/EIA/IS-95, *Mobile Station-Base Station Compatibility for Dual-Mode Wideband Spread Spectrum Cellular System*, Telecommunication Industry Association, 1993.

[10] S. Nanda, K. Balachandran, and S. Kumar, "Adaptation techniques in wireless packet data services," *IEEE Communication Magazine*, vol. 38, pp. 54–64, Jan. 2000.

[11] TIA/EIA/IS-707, *Data Service Options for Wideband Spread Spectrum Systems*, Telecommunication Industry Association, 1998.

[12] S. Ramakrishna and J. Holtzman, "Interaction of tcp and data access control in an integrated voice/data cdma system," *Mobile Networks & Applications*, vol. 3, pp. 409 – 417, Jan 1999.

[13] J. Herdtner and E. Chong, "Analysis of a Class of Asynchronous Power Control Algorithms for Cellular Wireless Systems," *IEEE Journal on Selected Areas in Communications*, vol. 18, pp. 436 – 446, March 2000.

[14] L. Song, N. Mandayam, and Z. Gajic, "Analysis of an Up/Down Power Control Algorithm for CDMA Reverse Link: A Nonlinear Control System Approach," *IEEE Journal on Selected Areas in Communications*, vol. 19, pp. 277–286, Feb. 2001.

[15] S. Kumar, S. Nanda, and A. Falconi, "Reverse Link Power Control Performance For High Speed Data Over CDMA," TR45.5.4/97.11,TIA Contribution, 1997.

[16] Y. Masuda and S. Whang, "Dynamic pricing for network service: Equilibrium and stability," *Management Science*, vol. 45, pp. 857 – 869, 1999.

[17] T. Abdelzaher and K. Shin, "End-host Architecture for QoS-Adaptive Communication," in *Real-Time Technology and Applications - Proceedings 1998. IEEE Comp Soc*, Los Alamitos, CA, 1998, pp. 121 – 130.

[18] D. Goodman and N. Mandayam, "Power control for wireless data," *IEEE Personal Communications Magazine*, vol. 7, pp. 48–54, April 2000.

[19] N. Feng, N. Mandayam, and D. Goodman, "Joint Power and Rate Optimization for Wireless Data Services Based on Utility Functions," in *CISS'99*, March 1999.

[20] C. Saraydar, N. Mandayam, and D. Goodman, "Efficient power control via pricing in wireless data networks," *IEEE Transactions on Communication*, To appear 2000.

[21] Viterbi, *CDMA: Principles of Spread Spectrum Communications*, Addison-Wesley Publishing Company, 1995.

[22] Release 99 TS25.211 215, *Physical Layer Standard*, UMTS 3GPP, 2000.

[23] D. Bertsekas, *Constrained Optimization and Lagarange Multiplier Methods*, Academic Press, 1982.

[24] M. Singh, *Dynamical Hierarchical Control*, North-Holland, New York, 1980.

[25] M. Zorzi, R. Rao, and L. Milstein, "Error statistics in data transmission over fading channels," *IEEE Transactions on Communication*, vol. 46, pp. 1468–1477, Nov. 1998.

[26] V. Garg and E. Sneed, "Digital wireless local loop system," *IEEE Communications Magazine*, pp. 112–115, Oct. 1996.

[27] T. Rockafellar, *Convex Analysis*, Princeton Univ. Press, Princeton, New Jersey, 1970.