Fair and Efficient Resource Allocation Model in a Wireless Multimedia CDMA System

Byung-Suk Min*, Yung-Tae Lee**, Jae-Hoon Kim*, Dong-Wan Tcha*

*Graduate school of management, Korea Advanced Institute of Science and Technology, 207-43 Cheongryangri-Dong, Dongdaemun-Gu, Seoul, 130-012, Korea

** Access Network Laboratory, Korea Telecom, 17 Woomyun-dong, Seocho-Gu, Seoul, Korea

ABSTRACT

In this paper, we address the resource allocation problem in wireless multimedia CDMA systems. The cellular CDMA system is interference-limited; the limit on the capacity is reached when anyone of SIR(signal to interference ratio) of each user is less than the predetermined threshold value. Given the minimum QoS(measured by bit energy to noise density), service rate(by bit per second) requirement and transmit power limit, we formulate the feasibility condition in terms of SIR instead of the transmitted power, which is mostly used in literatures, so that many advantages are obtained; smaller number of variables, simpler structure of feasible region, etc. From the perspectives of micro-economics, we discuss existing optimization criteria such as 'the minimum total transmitted power' or 'maximum sum of the rates,' which are to be verified not to guarantee 'fairness,' Accordingly, we suggest the new allocating decision criteria taking 'fairness' as well as 'efficiency' into account.

KEYWORDS
CDMA, resource allocation, SIR(signal to interference ratio)

1. Introduction

Direct sequence CDMA techniques have been extensive attention as an alternative wireless technology for the IMT-2000 system. The cellular CDMA system is interference-limited; the limit on the capacity is reached when anyone of SIR(signal to interference ratio) of each user is less than the predetermined threshold value. For spread spectrum systems, power control is necessary to achieve target quality of service. In a CDMA system with users sharing the same radio bandwidth and using the same radio site (base station) in each coverage area, inter-user interference adds on a power basis and the performance of each user becomes poorer as the number of simultaneous users increases. Power control of each user is essential to minimize each user's interference to the communication paths of other user.

For voice only CDMA systems, the transmitter power is to be changed to satisfy fixed rate and fixed QoS requirement. Thus, transmitted power should be controlled to maintain a certain signal to interference ratio which guarantee the given rate and QoS requirements. On the other hand, for multimedia CDMA systems, the rate requirement and QoS requirement is not equal from each other service. Moreover, the rate is variable, so that for more bandwidth or QoS, transmitted power should be increased to improve the SIR ratio.

In this paper, we address the problem of power control and resource management as a constrained optimization problem in the SIR, instead of power and rates. The constraints are specified in terms of the minimum QoS required, the maximum power that can be used and the minimum acceptable transmission rate. Different rates are accommodated by using different processing gains, while keeping the chip rate fixed. FEC coding is not explicitly considered. Capacity in terms of the number of users of each class that can be supported simultaneously, is evaluated.

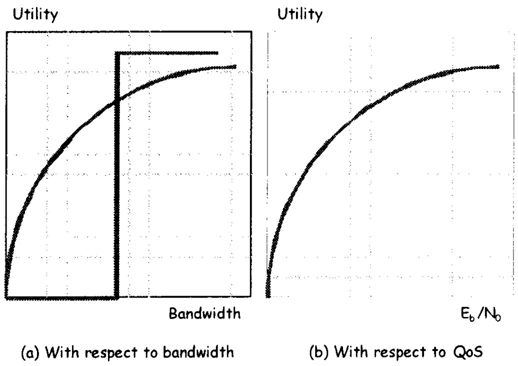
We consider the two-party model; system and users. The system allocates powers to each user, as a consequence, a user choose the combination of bandwidth and quality. In section 2 the system model is described and the choice problem of user is addressed. In section 3 two optimization criteria are solved for. The capacity of the system is also evaluate.

2. Optimal choice with utility factors: user

Concept of utility

What is the utility? The utility of a user is determined by qualitative factor and quantitative factor. In this work, service rate and QoS are considered as utility factors.

(1) Bandwidth. Let us discuss the qualitative properties of the application utility functions. We proceed by first holding BER constant so that utility only depends on delivered application bandwidth. For all applications, the application utility is a monotone non-decreasing function with respect to the bandwidth. We can categorize applications into many classes; nevertheless, we will discuss and contrast only two such classes. One class includes applications for which performance gradually improves as their allocated bandwidth increases, however with a decreasing marginal utility (e.g., video and text/graphics). The utility function for this class of applications is, therefore, concave everywhere, as shown as blue line in Fig. 1(a). Another class includes applications such as control information for which the received data are of no value to users if only partial information is delivered; however, once the necessary amount of data is delivered, there is no extra benefit for receiving more data. Red line in Fig. 1(a) shows the utility function for this class of applications.



<Figure 1> utility function with respect to the bandwidth and QoS

(2) Error rate. Let us now turn our attention to the other parameter of the utility function, that is, the error rate. When the received BER is high, users are generally unsatisfied with application performance. As the error rate improves, their satisfactions rise as well. However, once the BER improves beyond a certain level, very little additional satisfaction is achieved. For instance, the reception quality of video is nearly identical between BER's of 10^{-4} and 10^{-8}. Fig. 1(b) illustrates the utility function with respect to QoS, which is expressed by Eb/N0. (We assume that the relationship between Eb/N0 and BER is available.)

Developing a utility function

Let us introduce and consider a truncated root function as a utility function in the rest of his paper, which has a few realistic characteristics; first, a root function is monotone non-decreasing, second, the marginal utility decreases. Thus the utility function is

$$U = u(R,\gamma) = u(R)\cdot u(\gamma) = \sqrt{R\cdot\gamma}\text{ , if } R \geq r, \gamma \geq \delta \text{ ; 0, otherwise.}$$

Unfortunately, the truncated root function has unrealistic feature; the utility increases infinitely, in spite of the decreasing marginal utility, as long as the bandwidth or QoS increases. First of all, the maximum bandwidth is limited. Secondly, with respect to the error, the utility rather converges to some upper bound, than diffuses as root function. Ideally, with respect to bandwidth, non-decreasing function with the marginal utility decreasing and the double-sided truncated will do; with respect to QoS, converging function to a predetermined upper bound will do.

Budget curve

As a result of a power allocation, each user is given the relative signal power to the interference, SIR. Given the SIR, a user will select optimal rate/QoS combination from the SIR level according to his utility function.

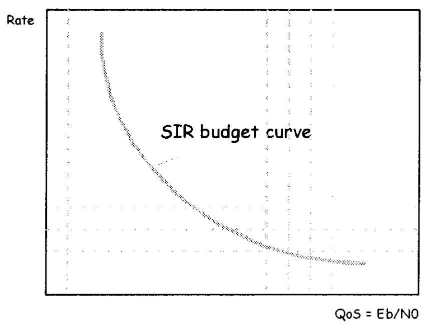
The shared frequency bandwidth is given as W, and uplink channel of a single cell system is considered. We assume that there are N users in the cell. The significant notations are as follows.

¶ $P = [P_1, P_2, \dots, P_N]$: power vector.
¶ $R = [R_1, R_2, \dots, R_N]$: rate vector.
¶ $G= [?_1, ?_2, \dots, ?_N]$: required E_b/N_0 vector.
¶ $p = [p_1, p_2, \dots, p_N]$: power upper bound (UB) vector.
¶ $r = [r_1, r_2, \dots, r_N]$: rate lower bound (LB) vector.
¶ $h = [h_1, h_2, \dots, h_N]$: channel gain vector.

The E_b/N_0 of each user is generally expressed by

$$\left(\frac{E_b}{N_0}\right)_i = \frac{W}{R_i}\cdot\frac{h_i P_i}{\sum_{j\neq i} h_j P_j + \eta_0 W}$$

From the perspective of a user, the power allocation is given by the system. Therefore he confronts the resource budget imposed by the power allocation. The budget curve is not linear because the product value of bandwidth and QoS is limited by the given SIR, on the contrary to the case of microeconomics, where the budget constraint is represented as a line. A budget curve is depicted in Figure 2.



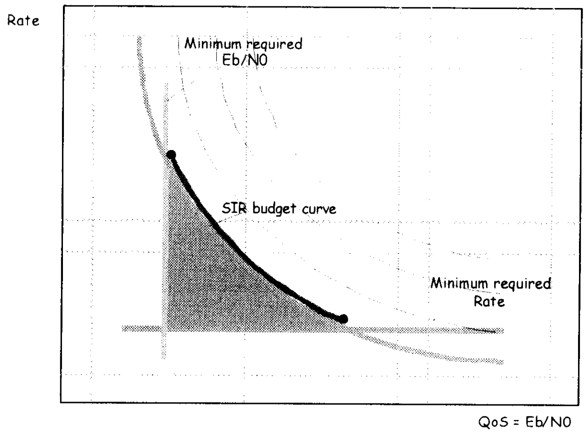
<Figure 2> budget curve of a user and feasible region

Optimal choice between the quantity and the quality

Under the SIR constraint, a user will try to maximize his utility by selecting the best combination of Eb/N0 and rate. According to the theory of consumer's choice, the optimal choice is the point where the slope of indifferent curve is same as the slope of budget curve. However, the utility curve exactly coincides the indifferent curve, optimal choices arise in every point on the efficient frontier.

$$\textbf{(P1)} \qquad \max_{R,\gamma} \; u_i(R_i,\gamma_i) = \sqrt{R_i\gamma_i}$$
$$\text{s.t.} \quad R_i\gamma_i \le W\lambda_i,$$
$$r_i \le R_i,\ \underline{\gamma_i} \le \gamma_i,$$
$$\text{where } \gamma_i = \left(\frac{E_b}{N_0}\right)_i,\ \lambda_i = \frac{h_i P_i}{\sum_{j\neq i} h_j P_j + \eta_0 W}.$$

The efficient frontier is depicted in a thick curve in Figure 4, in which the truncation of utility function is expressed by the constraint.



<Figure 3> Optimal choices of a user

Discussion. While, in above arguments, every points on the frontier is optimum all the solution, if one utility factor is more critical than the other one, the indifferent curve is distorted; if the quality (quantity) is more critical, the indifferent curve become gentler (steeper) slope so that the corner solution on the y-axis (x-axis) is obtained.
All things to be considered, the conclusion of the individual-scale analysis can be expressed by the words 'the more SIR, the more utility but the marginal utility decrease.'

3.Optimal allocation of resources: system

Sampath([2]) consider the following resource allocation feasibility conditions; ? power consumption constraint, ? minimum service rate constraint, and ? QOS guarantee constraint. Taking them into account, they formulate the feasibility region as follows.

$$\frac{W}{R_i} \cdot \frac{h_i P_i}{\sum_{j\neq i} h_j P_j + \eta_0 W} \ge \gamma_i,\ i = 1,\ldots,N \qquad (1)$$

$$0 \le P_i \le p_i,\ i = 1,\ldots,N \qquad (2)$$

$$r_i \le R_i,\ i = 1,\dots, N \tag{3}$$

However, this model has many disadvantages; first the number of variables is so many; second, the feasibility region is so complicated as to be hard to be analyzed, so that [Sampath] suggest a heuristic method in the throughput-maximization case.

Transformation and new formulation

Let us introduce the N to N nonlinear transformation; from the power variable into the SIR. With this transform we can reformulate the original resource allocation model.

$$\lambda_i = f_i(P_1, P_2, \dots, P_N) = \frac{h_i P_i}{\sum_{j \ne i} h_j P_j + \eta_0 W},\ i = 1,\dots, N$$

$$P_i = g_i(\lambda_1, \lambda_2, \dots, \lambda_N) = \frac{\eta_0 W}{h_i \cdot (1/\lambda_i + 1) \cdot \left(1 - \sum_j \frac{1}{(1/\lambda_j + 1)}\right)},\ i = 1,\dots, N$$

Theorem 1. The n functions are functionally independent; that is, the transformation is a one-to-one correspondence.
Proof. It is obvious from the fact that the Jacobian determinant is not equal to zero. ¦

The correspondences between the original constraints and the transformed ones are summarized in Table. 1.

[Table. 1] Correspondence between the original constraints and the transformed ones

	Original	Transformed
Power Non-negativity	$0 \le P_i,\ i = 1,\dots, N$	$\sum_{i=1}^{N} \frac{1}{1/\lambda_i + 1} < 1$
Rate constraints	$r_i \le R_i,\ i = 1,\dots, N$	$\underline{\lambda}_i \le \lambda_i,\ i = 1,\dots, N$
QoS constraints	$(E_b / N_0)_i \ge \gamma_i,\ i = 1,\dots, N$	
Power Constraint	$P_i \le p_i,\ i = 1,\dots, N$	$\sum_{i=1}^{N} \frac{1}{1/\lambda_i + 1} \le 1 - \frac{\eta}{\min_i Q_i (1/\lambda_i + 1)}$

In the following three theorems, we will verify the fact that given the objective functions, we can find another linear objective function, which has the same objective value under the feasibility conditions.

Claim 2. For the power minimization problem, the following two objectives has the same solution in respective space if necessary condition is satisfied.

$$\min \sum_{i=1}^{N} P_i \ ? \ \min \sum_{i=1}^{N} \lambda_i$$

Claim 3. For the throughput maximization problem, the following two objectives has the same solution in respective space if necessary condition is satisfied.

$$\max \sum_{i=1}^{N} R_i \ ? \ \max \sum_{i=1}^{N} \lambda_i$$

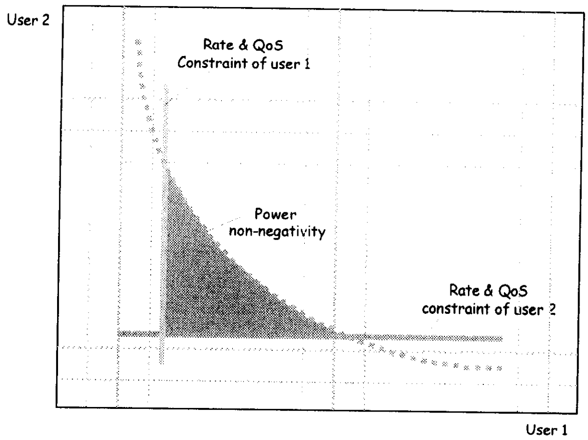
Claim 4. For the utility maximization problem, the following two objectives has the same solution in respective space if necessary condition is satisfied.

$$\min \sum_{i=1}^{N} \sqrt{R_i \gamma_i} \ ? \quad \min \sum_{i=1}^{N} \sqrt{\lambda_i}$$

The same feasibility conditions are simply formulated as follows. For the sake of explanation, we omit the power constraints.

$$\sum_{i=1}^{N} \frac{1}{1/\lambda_i + 1} < 1$$
$$\delta_i \leq \lambda_i$$

This inequality system has many advantages; first, the number of variables reduces to the half; second, the feasibility region becomes simpler structure to be analyzed. Now, let us consider the characteristics of the feasibility region. Fig. 4 illustrates the two-user system.



<Figure 4> A feasibility region : 2 user example

Decision criteria : old ones

The feasibility region is neither convex nor concave. This cannot be easily analyzed by conventional mathematical programming methods. Let us discuss the optimization criteria provided by [Sampath]; ? minimize the total transmitted power (so, power surplus minimization), ? maximize the throughput.

First, we can find optimal solution of the former objective is 'A' in the Fig. 5. (in this Figure, minimum required SIR is the same over all users, and the weight of each user is the same, which can be relaxed without difficulty.) The solution can be translated to be 'grudging allocation' such that no power surplus (SIR surplus) is allowed.

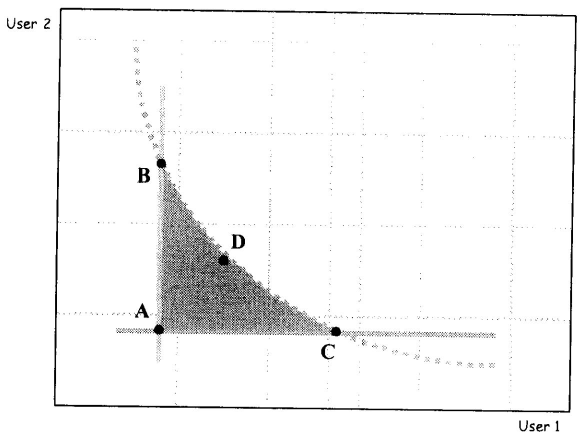
Next, we can find optimal solution of the latter objective is 'B' or 'C' in the Fig. 5. (for the simple explanation, we assume that the lower bound of lambda is the same.) The solution 'B' (respectively 'C') can be translated to be ' allocation' such that the power surplus is monopolized by user 2. (user 1) If we consider priority such as differentiated service classes composed of premium service user, best effort service user, etc, the tie would be broken by selecting the more weighted user.

How about the solution 'D'? From the solution, we can increase the throughput by giving the whole amount of surplus power to either user 1 or user 2, so that this is not optimal solution from the throughput-maximizing view.

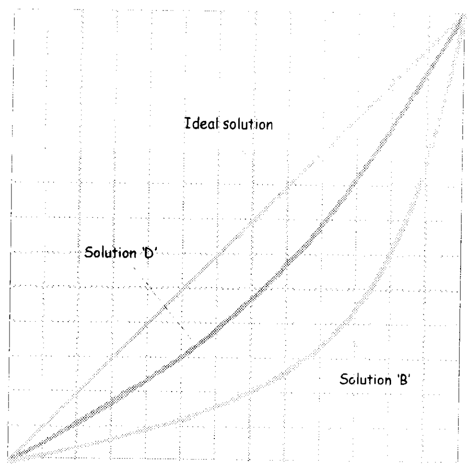
Thus it is not an 'efficient' solution. However, in the solution 'D', user 1 and user 2 share the surplus power while one user exclusively occupies it in the 'B' or 'C.' Therefore, it is a 'fair' solution.

Decision criteria : new one

In the theory of micro-economics, the fairness of the allocation is expressed by 'Lorenz curve' and measured by 'Gini coefficient.' For detailed introduction for Lorenz curve and Gini coefficient, see [1]. The Lorenz curves of the solutions 'B', 'D' are depicted in the Fig. 6.



<Figure. 5> Optimal solution from the criteria



<Figure 6> Lorenz curves of 'B' and 'D'

4. Remarks

In this paper, we have addressed issues in the support of multimedia services in a wireless CDMA system. In particular, we focused on optimality in SIR allocation from the system's view and in rate & QoS from a user's view. Different services will have different quality requirements and different resource (rate/power) constraints.

The problem was formulated as a constrained optimization of an objective function in the SIR. In particular, we focused on the minimum total transmitted power and the maximum total rate, and compared those with the new criterion, 'the maximum utility' and 'the fairness.' The former of existing ones tries to minimize the total interference caused by the current cell to other cells, while the latter tries to maximize throughput for users in the current cell.

We have developed closed form solutions for the three former cased; total power minimization problem, maximum total rate, maximum weighted utility. It was, however, assumed that the channel gains for all mobiles to their communicating base were known, a quantity that can usually be measured though the pilot channel. Thus, a centralized scheme is assumed. For the last, an efficient line search method was used, because of the difficulty of measuring the 'fairness.' Nevertheless, it was found to converge rapidly to the true solution.

A relevant extension of this work would be to incorporate the effect of delay into the optimization model as well as obtain results for the multi-cell case. The delay is, in fact, considered as the complementary utility factor to the QoS under the fixed bandwidth; wherein service is classified into two classes, data and voice. Data is delay tolerant but error-intolerant, conversely voice is error sensitive but delay intolerant.

5. REFERENCES

1. Hal Varian, Intermediate microeconomics.
2. A.Sampath et al., Power control and resource management for a multimedia CDMA wireless system, IEEE PIMRC, 1995.
3. Y.Lu and R.W.Broderson, Integrating power control, error correction coding, and scheduling for a CDMA downlink system, IEEE JSAC 17(5), 1999.