A MICRO-ECONOMICS APPROACH FOR SCHEDULING IN CDMA NETWORKS WITH END-TO-END QOS GUARANTEES

Walid Saad
American University of Beirut, Department of Electrical and Computer Engineering
Beirut, Lebanon, Email: wes02@aub.edu.lb

Sanaa Sharafeddine
Lebanese American University, Division of Computer Science and Mathematics
Beirut, Lebanon, Email: sanaa.sharafeddine@lau.edu.lb

Zaher Dawy
American University of Beirut, Department of Electrical and Computer Engineering
Beirut, Lebanon, Email: zaher.dawy@aub.edu.lb

ABSTRACT

Third generation CDMA networks strive to deliver high speed data services through a shared radio channel with scarce resources. To efficiently utilize the available radio resources, we propose a new scheduling algorithm based on techniques from micro-economics. Unlike existing literature that mainly focuses on maximizing total system and/or individual user utility, this new algorithm aims at ensuring QoS guarantees from end to end for all active connections. Moreover, it considers the time varying channel conditions in both uplink and downlink directions jointly rather than each direction separately. Simulation results show that the proposed algorithm allows several users to simultaneously transmit while providing end-to-end QoS guarantees in terms of frame success rate and end-to-end delay.

I. INTRODUCTION

The main driver behind the deployment of next generation CDMA-based cellular networks is their ability to provide mobile stations (MSs) with innovative multimedia services that demand high resources and impose stringent QoS requirements. One way to efficiently exploit the system resources is to use advanced scheduling algorithms. This resource allocation problem has been recently a major concern for a number of research papers; the majority of which sought to find the best scheduling algorithm in either the uplink or the downlink.

This paper proposes a novel scheduling algorithm for CDMA networks. The key distinctive features of this algorithm are three-fold: i. it schedules the users in communicating pairs based on the time-varying conditions of both the uplink and the downlink channels jointly, ii. its main allocation criterion is to provide end-to-end QoS guarantees, and iii. it is based on techniques taken from micro-economics such as indifference curves and marginal rate of substitution.

The paper is organized as follows. Section II presents a survey of related literature. Section III describes the system model and Section IV presents the proposed scheduling algorithm. Simulation results are presented and analyzed in Section V. Finally, conclusions are drawn in Section VI.

Early work in the area of uplink scheduling was done in [1] for a network having only voice users. The authors provide a power allocation algorithm that minimizes the total uplink transmitted power ensuring that every user achieves its target SNR. The authors in [2] and [3] propose an uplink scheduling algorithm for a CDMA network having a single data service. The uplink power control game is modelled as a non-cooperative game where every MS tries to maximize its utility function. Also, jointly with the scheduling algorithm, [3] considers a base station (BS) assignment algorithm that helps in meeting the objective of maximizing individual utilities.

Several next generation services are asymmetric and highly resource demanding in the downlink direction. Hence, exhaustive research activities focused on finding resource allocation algorithms suitable for the downlink. For instance, a joint BS and resource allocation algorithm is presented in [4] that maximizes the total utility (with price) per individual BS. This paper borrows the utility function used for the uplink in [3] and applies it to the downlink. The BS assignment scheme is modelled as a non cooperative game, and the output of the pricing based power and rate allocation problem is used in order to assign some MSs to highly loaded BSs to other lightly loaded BSs. In [5] and [6] the downlink resource allocation problem for a multi-cell heterogenous CDMA system is considered. In contrast to previous work, both papers take into account the rather non-critical delay requirements of these services. The power allocation problem is modelled using the Multiple Choice Multi Dimensional Knapsack (MMKP) problem. The main goal is to jointly allocate power and assign users to different BSs in a manner that maximizes total system utility.

The research done in [7] builds on the works of [5] and [6] in order to provide a new approach to the scheduling problem. The proposed algorithm allocates the total available transmit power of the BS to one MS during the transmission of a frame within a time slot. Moreover, unlike previous work, the utility function is defined per packet and not per user. The authors in [8] make yet another contribution in the downlink resource allocation problem. The aim of the paper is to find a joint BS assignment and power allocation algorithm that maximizes throughput and fairness in a CDMA system.
The majority of the reviewed papers focused on finding efficient scheduling algorithms that maximize the total network utility for either uplink or downlink users. Nevertheless, many next generation person-to-person services like video sharing or push-to-talk require end-to-end QoS between two MSs: an uplink MS sending packets to a downlink MS. In such end-to-end scenarios, the reviewed unidirectional schedulers might waste resources by allocating power to an uplink MS having excellent channel conditions but whose downlink recipient has a bad channel or vice versa. Hence, a more efficient end-to-end scheduler that can assess the peer to peer communication between two communicating MSs must be sought. In fact, the main motivation behind this work is to devise a new end-to-end scheduling algorithm that considers channel quality of the whole communication connection for MS-to-MS services (i.e. the joint channel quality of the uplink and downlink users).

III. SYSTEM MODEL

A. Network model

Consider a time-slotted multi-cell CDMA cellular system having $B$ base stations with omni-directional antennas and a single data service. The set of all BSs in the network is denoted by $A = \{a_n \mid n = 1 \ldots B\}$. A “connection” is defined as an uplink MS transmitting packets to a recipient downlink MS in an end-to-end session. Moreover, an “active connection” will refer to a connection that is scheduled for transmission. The network will consist of $2M$ MSs yielding a total of $M$ connections. Every user in the network will be assigned to the BS with the least path loss. The considered path loss model includes both distance based path loss and shadowing.

B. Signal To Noise Ratio

For every connection $l = (i,j)$ between an uplink MS $i$ sending data to a downlink MS $j$, the SNR of MS $i$ received by its serving BS $a_m$ in the uplink is given by

$$\Gamma_{l,i,a_m} = \frac{W}{R_i \sum_{k=1}^{N_{a_m}} g_{k,a_m} P_k + \alpha \sum_{k=1}^{N_{a_m}} g_{k,a_m} P_k + \sigma^2}$$

where $W$ is the chip rate, $R_i$ is the data rate of MS $i$, $P_k$ is the transmit power of MS $k$, $N_{a_m}$ is the number of MSs served by BS $a_m$, $\alpha$ is the intercell to intracell interference factor, $g_{i,a_m}$ is the path loss attenuation between MS $i$ and BS $a_m$, and $\sigma^2$ is Gaussian noise.

Similarly, the received SNR of MS $j$ served by BS $a_n$ in the downlink is given by

$$\gamma_{l,j,a_n} = \frac{W}{R_j \lambda g_{j,a_n} (P_{n,a} - P_j) + \alpha_j g_{j,a_n} P_{n,a} + \sigma^2}$$

where $R_j$ is the data rate of MS $j$, $P_j$ is the transmit power allocated to MS $j$ by BS $a_n$, $P_{n,a}$ is the total transmit power of BS $a_n$, $\lambda$ is the downlink orthogonality factor, $\alpha_j$ is the intercell to intracell interference factor of MS $j$, and $g_{j,a_n}$ is the path loss attenuation between MS $j$ and BS $a_n$. Since $P_{n,a}$ is typically much larger than $P_j$, the difference $P_{n,a} - P_j$ is assumed equal to $P_{n,a}$. Moreover, the intercell to intracell interference factor $\alpha_j$ is set equal to an average value for all MSs in the network.

The maximum total power of all BSs in the network is equal to $P_{BS}$. Therefore, for a BS $a_n$, the downlink maximum transmit power constraint can be written as

$$\sum_{k=1}^{N_{a_n}} P_k \leq P_{BS}$$

where $N_{a_n}$ is the number of MSs assigned to BS $a_n$. Similarly, the uplink transmit power of any MS $i$ is upper bounded by $P_{MS}$ as follows: $P_i \leq P_{MS}$.

C. Utility Function

For every connection $l$ in the network, a utility function $u_l$ is defined. The utility is a function of the time varying channel conditions as well as the end-to-end delay requirement of the service. This utility function will characterize every connection and will act as a QoS indicator. In order to account for the channel conditions, the frame success rate (FSR) is used as a channel quality indicator. For a particular direction (uplink or downlink), the FSR depends mainly on the modulation and coding schemes and is typically a function of the BER and, thus, a function of the SNR.

For the transmission of a packet from uplink MS to a downlink MS in a connection, the product of two FSR functions (one for the downlink and one for the uplink) will indicate the end-to-end FSR for delivering the packet from the sender to the receiver. This product is an increasing function of both uplink and downlink SNRs. Mathematically, the end-to-end FSR will be defined as follows:

$$f(\Gamma_{l,i,a_m}) f(\gamma_{l,j,a_n})$$

where $\Gamma_{l,i,a_m}$ and $\gamma_{l,j,a_n}$ are given by (1) and (2), respectively. The utility function $u_l$ of a connection $l$ will thus be a function of the end-to-end FSR represented by (4).

An approximation of the FSR will be used by adopting the sigmoidal function used in [4] which is a function similar in shape to the FSR of a BPSK modulated transmission. Hence, the FSR of a BPSK signal will be measured using the following equation for a MS $i$ served by BS $a_m$:

$$f(\Gamma_{l,i,a_m}) = c_i \left( \frac{1}{1 + e^{-\alpha_i(\Gamma_{l,i,a_m} - h_i)}} - z_i \right)$$

where $c_i = \frac{1+e^{\alpha_i h_i}}{e^{\alpha_i h_i}}$ and $z_i = \frac{1}{1+e^{\alpha_i h_i}}$. The shape of this sigmoid varies depending on the values of the parameters $\alpha$ and $h$. The values $\alpha = 3$ and $h = 3.5$ are used as in [4].

The delay model is inspired from [5] and [6] extended to the end-to-end delay case. End-to-end delay is defined as the elapsed time between two successive transmissions of a given connection based on end-to-end scheduling. For a given connection $l$, the end-to-end delay is represented by the maximum end-to-end tolerable delay $\tau_l$ which is considered to be equal to the sum of the maximum tolerable delays for both the uplink and downlink directions. Additionally, we define $d_l$ as the
remaining tolerable delay for a connection $l$ as follows:

$$d_l = \begin{cases} \tau_l - qT_s & \text{if } qT_s < \tau_l \\ 0 & \text{otherwise} \end{cases}$$

(6)

where $T_s$ is the duration of a time slot and $q$ is the number of time slots elapsed since the last transmission of connection $l$.

The proposed utility function will be a decreasing function of $d_l$, in a way that when $d_l$ approaches zero the corresponding term in the utility function will override the channel related term represented by the product given in (4). In a time slot, the utility of a connection $l$ is given by

$$u_l(\Gamma_{i,a,m}, \gamma_{j,a,n}) = f(\Gamma_{i,a,m})f(\gamma_{j,a,n})e^{\frac{-l}{Tu+dw}}$$

(7)

The parameter $T_w$ is defined as a correcting factor in order to bound the exponential term in (7) when $d_l$ approaches zero (i.e. the connection cannot tolerate more delay). Finally, since the SNRs are a function of the powers $P_i$ and $P_j$ for the uplink and downlink, respectively, the utility is also a function of these powers.

### D. Problem Definition

To ensure end-to-end QoS guarantee, a threshold $C$ for the utility function given by (7) is defined. This threshold would therefore be the minimum value that the utility of a connection $l$ must achieve in order to guarantee its QoS requirements. In summary, a connection $l$ would guarantee its QoS if

$$u_l \geq C$$

(8)

In every scenario, the threshold $C$ is set to a constant value for all connections in the network. When choosing $C$, one must note that for a chosen value of $C$ the worst case/minimum required end-to-end FSR is $C/e^{\frac{t}{Tu+dw}}$. This case occurs when a connection has not transmitted for $\tau_l/T_s$ time slots, i.e. $d_l = 0$.

The end-to-end scheduling problem is defined as follows:

**Problem.** In every time slot, find the power tuples $(P_i, P_j)$ that must be allocated for every connection $l = (i, j)$ in a way to maximize the number of connections in the network satisfying (8). These connections will consequently be allowed to transmit in the considered time slot. A connection that will not be active in the considered time slot will be assigned the power tuple $(P_i, P_j) = (0, 0)$.

### IV. The Scheduling Algorithm

#### A. Marginal Rate of Substitution

In microeconomics, an indifference curve is a plot representing different combinations of two goods which yield the same utility value. For example, for a consumer who wants to buy $x$ apples and $y$ bananas, the utility function $u(x,y)$ represents the calories that the consumer gains from the $x$ apples and $y$ bananas. The indifference curve represents the various combinations of apples and bananas that give the consumer a constant number of calories (i.e. utility value) $U$. Mathematically, the indifference curve can be defined as $u(x,y) = U$ whereby the x-axis represents the quantity of apples (good 1) and the y-axis the quantity of bananas (good 2). An example indifference curve is shown in Figure 1.

The Marginal Rate of Substitution (MRS) at any point of an indifference curve is defined as the rate at which a consumer is willing to substitute one good for the other. Substituting one good for another means that the consumer will move from one point of the indifference curve to another. For example, in Figure 1 a consumer might choose to decrease the quantity of good 2 by $\delta y = y_2 - y_1$ and increase the quantity of good 1 by $\delta x = x_2 - x_1$ while maintaining the same utility value $U$. In this case, the consumer would be moving from point $(x_1, y_1)$ to point $(x_2, y_2)$ on the indifference curve. In such a case, the following computations would yield the MRS:

$$u(x, y) = U$$

(9)

By differentiating we find $du = 0$, thus

$$(\partial u/\partial x)\delta x + (\partial u/\partial y)\delta y = 0$$

(10)

Through equation (10) the MRS will be defined as

$$\omega_{x,y} = -\frac{\delta y}{\delta x} = \frac{(\partial u/\partial x)}{(\partial u/\partial y)}$$

(11)

Note that $\delta x$ and $\delta y$ are always of opposite signs since on an indifference curve a decrease of the quantity of one good (e.g. negative $\delta y$ for good 2) always yields an increase in the quantity of the other good (e.g. positive $\delta x$ for good 1).

For every connection, the utility function defined by (7) is a function of two variables: the uplink power $P_i$ and downlink power $P_j$. The following equation represents the minimum acceptable QoS constraint presented by (8):

$$u_l(P_i, P_j) = C$$

(12)

By comparing with (9), (12) is clearly the equation of an indifference curve with the $x$ and $y$ axes (quantities of good 1 and good 2) representing the uplink and the downlink power values $P_i$ and $P_j$, respectively. In fact, this indifference curve represents the different combinations of uplink and downlink powers that yield a constant utility value equal to $C$. 

Figure 1: An example indifference curve.
From (12), one can conclude that for a connection \( l = (i, j) \) two main contributors in the end-to-end utility value exist: the uplink MS transmit power \( P_i \) and the downlink BS transmit power \( P_j \). In order to distribute this contribution fairly between these two contributors, it would be of interest to operate the connection at the point of the indifference curve where the MRS \( \omega_{P_i, P_j} \) is one (i.e. \( \delta P_i = -\delta P_j \)), since it would yield a fair contribution in terms of uplink and downlink powers to the total utility. Operating at points where \( \omega_{P_i, P_j} \neq 1 \) means that moving in the negative or positive x-direction yields a saving in the contribution of one of the powers (uplink or downlink), thus justifying the choice of \( \omega_{P_i, P_j} = 1 \).

**B. MRS-based Algorithm**

The MRS-based solution to the end-to-end resource allocation problem would be to find for every connection \( l = (i, j) \) in the network the power tuple \( (P_i, P_j) \) that allows \( l \) to achieve a particular utility threshold value \( C \) while operating at \( \omega_{P_i, P_j} = 1 \). Mathematically this maps into solving the following system of two equations:

\[
\begin{align*}
    u_l(\Gamma_{i,a_m,\gamma_{j,a_n}}) &= C \quad (13) \\
    \omega_{P_i, P_j} &= 1 \quad (14)
\end{align*}
\]

Equations (11) and (14) yield:

\[
\frac{\partial u}{\partial P_i} = \frac{\partial u}{\partial P_j} \quad (15)
\]

A heuristic iterative approach is used for modeling the uplink intracell interference. For an uplink BS \( a_m \), the algorithm starts by assuming the intracell interference equal to the following:

\[
I_{\text{intra}} = \sum_{k=1}^{N_{a_m}} g_{k,a_m} \hat{P} \quad (16)
\]

where \( N_{a_m} \) is the number of uplink MSs assigned to BS \( a_m \). In every time slot, \( \hat{P} \) will be given different values in several iterations. For example, these iterations would start in a time slot with \( \hat{P} = 1 \) mW and increment \( \hat{P} \) by a specific step (e.g. 1 mW) in subsequent iterations. The number of iterations will be limited by the maximum uplink transmit power \( P_{\text{MS}} \).

The main driver behind using multiple iterations for modeling the uplink intracell interference is that each iteration aims at assuming an intracell interference value as close as possible to the actual value resulting from solving the system of equations (13) and (14). For a particular time slot, the MRS-based scheduling algorithm performs the steps specified by Algorithm 1 repeated in every time slot. Note that \( B_{a_i} \) and \( B_{a_i} \) denote the number of BSs serving users in the uplink and the downlink respectively.

In Algorithm 1, the sorting of connection groups in increasing power (for the uplink in Step 3) and decreasing power (for the downlink in Step 6) is performed in order to activate the largest possible number of connections. On one hand, Step 3 allows the algorithm to minimize the number of deactivated connections due to the iterative intracell interference approach. On the other hand, Step 6 allows the algorithm to admit a maximum number of connections satisfying the downlink power constraint given by (3).

**Algorithm 1 Proposed MRS-based scheduling algorithm.**

```
for \( \hat{P} = 0.001 \) W to \( P_{\text{MS}} \) [increments of 0.001 W] do
    for \( l = 1 \) to \( M \{ l = (i, j) \} \) do
        Step 1. Solve the system of equations (14) and (15) in order to obtain the tuple \( (P_i, P_j) \).
        if Solution found then
            Store power tuple \( (P_i, P_j) \)
        else
            Deactivate connection \( l \) by setting \( (P_i, P_j) = (0, 0) \)
        end if
    end for
step 2. Group the connections by groups served by the same uplink BS
    Step 3. Sort the connections in each group by decreasing order of the uplink power allocated \( P_i \).
    for \( b = 1 \) to \( B_{a_i} \) do
        Step 4. Recompute the resulting uplink intracell interference
        if resulting uplink intracell > assumed value then
            for all connections served by the UL BS of iteration \( b \) do
                Recompute the downlink power value \( P_j \) through equation (12) using the resulting uplink intracell interference
                if \( P_j \leq 0 \) or \( P_j > P_{\text{BS}} \) then
                    Deactivate the connection and update accordingly the resulting uplink intracell value for this BS
                end if
            end for
        end for
    end for
    Step 5. Group the connections by groups served by the same downlink BS
    Step 6. Sort the connections in each group by increasing order of the downlink power allocated \( P_j \).
    for \( p = 1 \) to \( B_{a_i} \) do
        while total power of activated connections < \( P_{\text{BS}} \) do
            The remaining connections at the end of the loop are deactivated by setting \( (P_i, P_j) = (0, 0) \) do
                Go through the sorted connections group served by the downlink BS of iteration \( p \) and activate them one by one by storing their power tuples \( (P_i, P_j) \)
            end while
        end for
    end for
    Step 7. Compute the number of active connections for this iteration of \( \hat{P} \) by counting the number of connections having \( (P_i, P_j) \neq (0, 0) \).
    end for
    Step 8. The iteration of \( \hat{P} \) that will give the maximum number of active connections is selected and the active connections will be allocated the computed power tuples \( (P_i, P_j) \) in that iteration.
```

**V. Simulation Results and Analysis**

The set of parameters describing the simulated cellular network are summarized in Table 1. For each simulation scenario, the threshold is chosen as \( C = 1 \) in order to ensure a high end-to-end FSR for any active connection. The simulation period is 30 sec (3000 time slots). Simulation statistics are shown in Table 2 that show the maximum, minimum and mean (over time) values for the percentage of active connections and the mean FSR (mean end-to-end FSR of all users in a time slot) recorded over the duration of the simulation. The simulation results show that the proposed algorithm is able to service an average of 72.08% of active connections over 30 seconds. In addition, the best performance achieved was maintaining 96.43%
of the connections in the network active in certain time slots while meeting their end-to-end QoS requirements. Moreover, the algorithm was able to guarantee for every connection in the network an end-to-end FSR of no less than 0.992.

As the total number $N_t$ of users per BS increases, Figure 2 shows that the average number of connections increases due to the increased multiuser diversity. Moreover, the algorithm was able to maintain a rather constant mean FSR of around 0.9992 over all cases. The final simulation considers a network scenario with delay sensitive connections having $\tau = 30$ ms.

### Table 3: Delay sensitive vs. delay tolerant network scenarios.

<table>
<thead>
<tr>
<th>$\tau$ (ms)</th>
<th>$T_w$</th>
<th>Mean FSR (over time)</th>
<th>% of active connections</th>
</tr>
</thead>
<tbody>
<tr>
<td>30</td>
<td>0.2</td>
<td>0.9983</td>
<td>79.24</td>
</tr>
<tr>
<td>300</td>
<td>2</td>
<td>0.9998</td>
<td>72.08</td>
</tr>
</tbody>
</table>

VI. C ONCLUSIONS

This paper proposed a micro-economics based algorithm that allows MSs in a CDMA network to transmit while maintaining end-to-end QoS guarantees. The algorithm schedules the users in communicating pairs allowing both the uplink and downlink channels to contribute in achieving a target end-to-end utility.

REFERENCES


