<table>
<thead>
<tr>
<th>Title</th>
<th>Adaptive packet scheduling in cellular CDMA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Author(s)</td>
<td>Zhuge, L; Cao, Y; Li, VOK</td>
</tr>
<tr>
<td>Citation</td>
<td>IEEE International Symposium on Personal, Indoor, and Mobile Radio Communications, PIMRC, 2003, v. 2, p. 1192-1196</td>
</tr>
<tr>
<td>Issued Date</td>
<td>2003</td>
</tr>
<tr>
<td>URL</td>
<td><a href="http://hdl.handle.net/10722/46492">http://hdl.handle.net/10722/46492</a></td>
</tr>
</tbody>
</table>

Rights: This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License. ©2003 IEEE. Personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution to servers or lists, or to reuse any copyrighted component of this work in other works must be obtained from the IEEE.
Adaptive Packet Scheduling in Cellular CDMA

Lei Zhuge
Department of Mathematics
University of Southern California
Los Angeles, CA 90089
Email: lzhuge@usc.edu

Yaxin Cao
Broadcom Corporation
1715 One Arin Park, Hwy 35
Middletown, NJ 07748
Email: yaxinc@broadcom.com

Victor O. K. Li
Department of Electrical and Electronic Engineering
The University of Hong Kong
Pokfulam, Hong Kong, China
Email: vli@eee.hku.hk

Abstract—An adaptive packet scheduling algorithm for cellular CDMA systems is proposed. The algorithm guarantees packet deadline and average data rate under the assumption of perfect power control. Channel condition is also considered to reduce the transmission power.

I. INTRODUCTION

Packet data services have been proposed for next generation mobile communication systems [1]. Some people have even mentioned all-IP wireless networks [2]. A typical problem used to be studied in wired networks — packet scheduling — has thus drawn much attention from wireless network researchers [3]. Compared with wired networks, wireless networks have several exclusive characteristics such as error-prone channel (with possibly bursty errors), varying channel capacity, and user mobility, etc. These issues must be handled properly when we design wireless packet scheduling algorithms.

Besides general wireless system characteristics, direct sequence code division multiple access (DS-CDMA) has a few specialties for the packet scheduling problem. Since in such systems several channels can be scheduled for simultaneous transmission, those packet scheduling algorithms designed for TDMA-based single-server systems (e.g. [4], [5]) are generally not applicable. On the other hand, variable service rates provided by multi-code CDMA [6] can be used to increase flexibility in packet scheduling. Furthermore, the fast closed-loop power control enables the base station to gather fairly good estimates of the channel condition, which facilitates the design of packet scheduling algorithms with channel status awareness.

In existing work on CDMA packet scheduling, [7] and [8] both designed slot-based algorithms, where a CDMA frame is divided into multiple slots for transmitting data and control packets to/from different users. While slotted-CDMA may achieve higher capacity from reducing intra-cell interference, it requires smart slot assignment algorithm (which is usually very difficult to design) and fundamental modifications of frame structure and media access control (MAC) protocol from the existing CDMA proposals.

In [9] packets are scheduled according to their calculated priorities and the system resource constraints. The priority of a packet is defined as inversely proportional to the remaining time before the packet expires. Our algorithm being proposed has a similar strategy to [9] in the sense that the CDMA frame structure and MAC protocol are maintained, and priority-based scheduling is performed. However, our packet scheduling algorithm is adaptive to not only the packet delay deadline required by the application, but also the deviation of data rate from its target value and the channel status variations. In the CDMA scenario the channel status is more coupled with power consumption than packet loss, unlike TDMA, which is not yet considered by any of the CDMA packet scheduling methods.

In addition, our algorithm works with realistic channel model instead of the simple two-state model used in the majority of TDMA packet scheduling methods (if at all they consider the issue). We will discuss the relations between packet scheduling and flow admission control as well.

In the next section we review the capacity constraints and packet transmission scenarios in cellular CDMA systems, including a few assumptions, for the following discussions on packet scheduling. The adaptive packet scheduling algorithm is presented in section III, where the relations between flow admission and packet scheduling are also discussed. We give some numerical examples of our algorithm in section IV. and conclude the paper in section V.

II. CDMA SYSTEM RELATED ISSUES

A. Capacity Constraints

In CDMA all streaming service (e.g., voice) and packet data service users share the same spectrum. In this paper we assume streaming services always have the highest priority and are one hundred percent guaranteed (i.e., no loss), and thus we have variable amount of remaining resource to be allocated by the packet scheduling algorithm for packet data users. (Note that our algorithm can be applied to all-packet-data systems as well.) The uplink (from mobile user to base station) capacity of a CDMA system is limited by the total interference power in the system, while in the downlink (from base station to mobile user) it is the maximum base station power that limits the system capacity. For convenience in this paper we consider a single-cell system (i.e. other-cell interference is not included).

For either uplink or downlink, we have found a capacity constraint relating the data rate $R_i$ and the received bit energy to interference density ratio (BIR) $\gamma_i = \frac{S_i}{I_i}$ of all user $i$ to the system channel bandwidth $W$, where $S_i$, $I_i$, and $R_i$ are received signal power, interference power, and data rate of user...
detected by the bit error rate requirement.

Specifically [10], [11], for the uplink of a single-cell system, the capacity constraint is

\[ \sum_{i=1}^{N} C_i \leq W(1 - \eta) \]  

(1)

where \( N \) is the number of users in the cell, \( \eta \) is the noise to interference density ratio (a quantity for the interference limit), and \( C_i = R_i \gamma_i \) or \( C_i = \left[ \frac{1}{R_i \gamma_i + \frac{1}{W(1-\eta)}} \right]^{-1} \) is called the virtual bandwidth consumed by user \( i \). The difference between the two expressions of \( C_i \) is: The simpler expression is an approximation where the total received signal power from all users is counted as the interference power, while the more complicated expression has excluded the desired user signal power from the interference power.

The capacity constraint for the downlink of a single-cell CDMA system is

\[ \sum_{i=1}^{N} C_i \leq \frac{W(1 - \beta)}{1 - f_o} \]  

(2)

where \( \beta \) is the fraction of the maximum base station power reserved for control purposes (synchronization, paging, pilot, etc.), \( f_o \) or the orthogonality factor is the fraction of interference (at the user side) cancelled by the downlink orthogonal codes, and \( C_i = R_i \gamma_i \) or \( C_i = \left[ \frac{1}{R_i \gamma_i + \frac{1}{W}} \right]^{-1} \). The reason for our having two expressions for \( C_i \) is exactly the same as in the uplink.

Since the different \( C_i \) expressions only slightly change the capacity calculation, but do not affect the packet scheduling algorithm, for simplicity we just use the simpler version \( C_i = R_i \gamma_i \) in this paper for both uplink and downlink.

B. Packet Transmission Scenarios

In the existing CDMA proposals data packets are transmitted in fixed-length frames (e.g. 10 ms). In general cases user transmission rates are variable, and thus the packet scheduling problem becomes transmission rate assignment problem (in the fixed-length frames) for the users requiring data services. Variable transmission rates can be realized by the multi-code CDMA scheme [6], where multiple packets from a user are transmitted at the same time on associated code channels.

In order for the base station to perform the packet scheduling or transmission rate assignment, the users need to send requests before data transmission. Sending the request for each data packet is inefficient. However (as in [7]), it is reasonable to assume that packets arrive in batches (bursts), and all the packets in a batch have the same delay deadline. Note that the packet delay deadline in this paper is not directly associated with the data rate, but is determined by the application requirement. A user sends to the base station a burst transmission request for all the packets in the burst. This request should include the amount of data (e.g. number of bits) in the burst and the delay deadline of the burst.

The base station needs to maintain a request queue for each user who has sent a burst transmission request. Prior to each frame time, the base station performs packet scheduling algorithm for the head-of-line burst transmission requests of all request queues, and assigns the data rate of transmission from each user in the next frame. From section II-A we know that when the required receiving BIR \( \gamma_i \) is given for each user, the cell capacity is actually a constraint on the transmission data rate \( R_i \) from each user, and thus it is a rate assignment problem.

Besides the issues of packet delay deadline and cell capacity, the packet scheduling algorithm may also need to control the average data rate from the user, and consider the channel status of each user (to avoid too much transmission power from the user with a bad channel). In the next section we propose a packet scheduling algorithm taking all these issues into consideration.

The downlink scenario is a little simpler since the base station has all the knowledge of the data to be transmitted. However, the base station may ask a user to report the pilot strength measurement before transmitting data to it in order for the base station to estimate the channel status and determine the transmission power. Assume the base station knows the delay allowed for transmitting each packet on this last hop to the receiver (from the delay information tagged on the packet, for example). The downlink packet scheduling algorithm should be very similar to the uplink one.

III. ADAPTIVE PACKET SCHEDULING ALGORITHM

Since the uplink and downlink have similar capacity expressions (eq. (1) and (2)) and packet scheduling procedures, we use uplink in presenting our packet scheduling algorithm.

Here is the overview of the algorithm: First a priority is calculated for the packets in the head-of-line bursts of each request queue according to their delay deadline and the data rate lag (will be defined) of each user. Then based on its channel status, the cell capacity, and also the data rate lag, a transmission rate for the next frame is assigned to each user following the order of their packet priorities.

A. Priority Calculation

The priority of each packet consists of two parts: delay guarantee priority and lag adjustment priority.

A packet with closer deadline should have a higher priority of transmission to avoid its loss due to expiration. Let \( t \) be the current time, \( a \) be the packet arrival time, and \( d \) be the maximum delay of the packet before expiration, the time-to-live of the packet is defined to be

\[ t_l = \frac{a + d - t}{T_f} \]  

(3)

where \( T_f \) is the frame time. In the request queue at the base station, the packet arrival time is approximated by the arrival time of the request so the time of generating and transmitting the request is ignored. (In the downlink we can have accurate packet arrival time.)
By its definition, \( t_i \) is actually the remaining time of the packet in the unit of frame time. The *delay guarantee priority* is thus chosen to be inversely proportional to \( t_i \).

On the other hand, a user with larger lag on the average data rate should have higher transmission priority, so that it can eventually reach the target data rate. Assume that the target data rate and BIR have been agreed on between the user and the base station during the flow admission phase at the beginning of their communication (see section III-C for details). The *data rate lag* is defined as the difference between the target data rate \( R^* \) and the average data rate \( \bar{R} \) from the beginning of the flow until the current time, or \( \Delta R = R^* - \bar{R} \).

If \( \Delta R < 0 \) which means the user has received more service than it needs up to now, its priority in the lag adjustment part should have negative value so that it gives the transmission opportunity to users with positive lags. Also the lag adjustment priority should be related to the *normalized lag* with respect to the target data rate.

We design the lag adjustment priority as (for each user)

\[
p_{\text{lag}} = \frac{\Delta R}{\mu R^* - |\Delta R|}, \quad |\Delta R| \leq \mu R^* \tag{4}
\]

where \( \mu \) is the maximum allowed deviation of the average data rate from the target value. Note that \( p_{\text{lag}} \) has the same sign as \( \Delta R \), and its absolute value increases faster when the data rate deviation \( |\Delta R| \) gets closer to its limit \( \mu R^* \) (can be seen from the first derivative of \( p_{\text{lag}} \)). In case \( |\Delta R| > \mu R^* \), just set \( |\Delta R| = \mu R^* \) in the calculation so that \( p_{\text{lag}} \) goes to 0 or \(-\infty\) depending on the sign of \( \Delta R \) (in implementation we need to specify a big but finite number for \( \infty \)).

Lag adjustment is also the way to achieving fairness among the users. We regard a user as fairly serviced if its target data rate is kept at its target value, fairness is achieved when the average data rate of each user reaches its target data rate.

Combining the delay guarantee priority and lag adjustment priority, the overall packet transmission priority is calculated as

\[
p_{R} = \alpha p_{\text{lag}} + \frac{T_f}{d-t}, \quad |\Delta R| \leq \mu R^* \tag{5}
\]

where the coefficient \( \alpha \) is configured to achieve a compromise between the lag adjustment and delay guarantee, or user fairness and data loss. Note that at a certain time all the packets in a head-of-line burst must have the same priority, since the delay deadline and data rate deviation are the same for all of them, so they can be all transmitted at a time if enough data rate is assigned to the corresponding user.

### B. Transmission Rate Assignment

Due to power control, the transmission power from a user is proportional to the channel degradation, or inversely proportional to the fading-times-shadowing envelope given the path loss. When channel status is unfavorable, the user needs large transmission power to combat fading and shadowing. If even the maximum power in the user device can not satisfy the transmission requirement, the transmitted packet will be received at a low BIR and thus a high probability of error. The packet scheduling algorithm, therefore, should take the channel status into consideration, and lower the user transmission power as much as possible.

For simplicity, in this paper we assume perfect power control which means the received BIR \( \gamma \), averaged over every frame is kept at the required value for each user. In our scheduling algorithm, when the channel degradation of a user is too serious (exceeding a predefined threshold), the user will not be allowed to transmit in the next frame, and the opportunity is given to other users with better channels. In this way the average transmission power from a user and the total transmission power from all users can both be reduced. The channel degradation threshold should be set such that it takes effect before the maximum power in the user device is reached. After all these considerations we can assume that packets are always received with the desired quality.

The user transmission power, which is proportional to the channel degradation, is equal to the transmission bit energy times the data rate. Therefore, to avoid large transmission power when channel degradation is serious, we should assign data rate as inversely proportional to channel degradation. In practice the base station can estimate the channel status from the user's *unit-rate transmission power* defined as the transmission power divided by the data rate (so that the data rate factor in the transmission power is eliminated and only the channel status effect is left). Since the base station issues all the power control commands instructing the user to increase or decrease transmission power, it should be able to track the change of the transmission power from the user. Suppose the initial unit-rate transmission power from a user is \( s_0 \) when the flow starts (\( s_0 \) may be unknown to the base station in the uplink1). Let \( s_t = g_0 s_0 \) be the current unit-rate transmission power from the user, and \( s = g_0 s_0 \) be its average unit-rate transmission power since the flow starts. The base station knows \( y_t \) and \( g \) because the power control commands actually specify relative power changes (e. g. increase 1 dB from the previous value). The ratio \( s_t/s = g_t/g \) is used in our algorithm as an index of the channel status (smaller \( s_t/s \) or \( g_t/g \) corresponds to better channel). We use this ratio over the average value in order to decouple the channel path loss from the transmission rate assignment as both \( s_t \) and \( s \) contain the same path loss component. Otherwise users closer to the base station tend to receive more services since their path losses are smaller, which is not a reasonable assignment scheme. *Transmission rate is thus assigned to be inversely proportional to \( s_t/s \) or \( g_t/g \).*

However, the current value \( s_t \) or \( g_t \) may not be a good estimate of the channel status for data transmission in the next frame, if we consider the fast channel variations (may be much

1In the downlink the base station obviously has full knowledge of the transmission power.
faster than the frame rate). A better estimate, as suggested in [12], is the "local average" of the channel envelope which is the average value over a few frames. Therefore, the $s_t$ or $g_t$ used in our algorithm is actually a value averaged over a time window $v_{int}$, which covers a fixed number of frames up to the current time. Note that if a user has no transmission in a frame, this frame time should not be counted in the calculation of the local average $s_t$ or $g_t$.

On the other hand, the assigned transmission rate should be proportional to the target value. In order to let the average data rate converge to the target value more quickly, the lag adjustment is also included in our transmission rate assignment. The final assigned transmission rate is (for each user)

$$R_t = (R^* + \Delta R) \left( \frac{\delta_t}{g_t} \right) - 1 = (R^* + \Delta R) \left( \frac{\delta}{g_t} \right),$$  \hspace{1cm} (6a)

where $\xi$ is the channel status threshold such that

$$R_t = 0 \quad \text{when} \quad \frac{\delta}{g_t} < \xi$$  \hspace{1cm} (6b)

In addition, we put a lower limit $R_{\text{min}}$ on $R_t$

$$R_t \geq R_{\text{min}} \quad \text{when} \quad \frac{\delta}{g_t} \geq \xi$$  \hspace{1cm} (6c)

because if a user has enough priority and the channel degradation is below the threshold it should be allowed to transmit a certain amount of backlogged data. $R_{\text{min}}$ also corresponds to the rate of the "fundamental channel" in multi-code CDMA. A user is allowed to send all the packets of a head-of-line burst in the next frame if a large enough $R_t$ is assigned by (6a).

C. Relations to Flow Admission

Like streaming services, packet data flows need to be admitted as well before requesting burst transmissions. In the flow admission phase users tell the base station about their target data rates and quality of services requirements (BIR). The base station admits flows based on its estimation on the resource required by each flow. The estimation may be characterized by the mean and variance of the resource consumption by each flow. Under the assumption of perfect power control and no other-cell interference, the only variation of the resource consumption (the virtual bandwidth $R_t$) comes from the data rate. The total virtual bandwidth $C = \sum C_i$ in this case can usually be approximated by a Gaussian random variable. Given the outage probability $\delta$ in the uplink, which is the probability of the total interference power exceeding the acceptable level, or equivalently the probability of $C > W(1 - \eta) \{\text{eq. (1)}\}$, the flow admission criterion for the uplink is [10]

$$m_v + \sigma_v \leq W(1 - \eta) \leq C$$  \hspace{1cm} (7)

where $\sigma = Q^{-1}(\delta)$ with $Q(s) = \frac{1}{\sqrt{2\pi}} \int_s^\infty e^{-t^2/2} dt$ being the tail probability function of a (0, 1) Gaussian random variable, and

$$m_v = \sum m_i \gamma_i, \quad \sigma_v = \sum \sigma_i \gamma_i^2$$  \hspace{1cm} (8)

are the mean and variance of the total virtual bandwidth $C$. The mean and variance of $R_t$ are $\mu_R$ and $\sigma_R$. -- can be either estimated from a known traffic model of the flow from user i or reported by the user. We can also write a similar flow admission constraint for the downlink.

If no packet scheduling is performed, the portion of data causing the total virtual bandwidth $C$ exceeding $W(1 - \eta)$ is regarded as being lost. Applying the Gaussian approximation, the average loss rate is, approximately,

$$P_l = E \left( \frac{C - W'}{C} \right) \geq W' \right) P_r \left( C > W' \right)$$

$$= \delta \int_{W'}^{\infty} \frac{1}{\sqrt{2\pi} \sigma_v} \left( \frac{C - W'}{C} \right) e^{-\left(\frac{C - W'}{\sigma_v}\right)^2} dc,$$  \hspace{1cm} (9)

where $\delta$ is the outage probability used in flow admission. The integral in this formula can be evaluated numerically when all the parameters are given. Packet scheduling can take advantage of the delay allowed for each packet and thus reduce the packet loss rate.

IV. Numerical Examples

We consider the uplink of a single-cell system with three types of traffic: voice, low speed (type I) data, and high speed (type II) data. The simulation parameters are listed in Table 1. The voice stream is modelled as an on-off source with exponentially distributed talk and silent spurts. The data traffic (both types) have Poisson-arrival packets. For simplicity of the simulation we define the same delay deadline for all packets in a data type. The channel model is the product of a Rayleigh fading process and a lognormal shadowing process.

<table>
<thead>
<tr>
<th>Item</th>
<th>Symbol</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>System bandwidth</td>
<td>$W$</td>
<td>2.5 MHz</td>
</tr>
<tr>
<td>Noise to interference ratio</td>
<td>$\eta$</td>
<td>0.1</td>
</tr>
<tr>
<td>Rayleigh fading parameter</td>
<td>$\sigma_\gamma$</td>
<td>0.5</td>
</tr>
<tr>
<td>Shadowing deviation</td>
<td>$\sigma_s$</td>
<td>6 dB</td>
</tr>
<tr>
<td>Lag adjustment priority coefficient</td>
<td>$\alpha$</td>
<td>1</td>
</tr>
<tr>
<td>Maximum lag deviation factor</td>
<td>$\mu$</td>
<td>0.1</td>
</tr>
<tr>
<td>Outage probability</td>
<td>$\delta$</td>
<td>0.1</td>
</tr>
<tr>
<td>Frame length</td>
<td>$T_f$</td>
<td>10 ms</td>
</tr>
<tr>
<td>Local average window</td>
<td>$w_{av}$</td>
<td>50 ms</td>
</tr>
<tr>
<td>Simulation length</td>
<td>$T_s$</td>
<td>50000 frames</td>
</tr>
<tr>
<td>Voice rate</td>
<td>$R_v$</td>
<td>14.4 Kbps</td>
</tr>
<tr>
<td>Voice activity factor</td>
<td>$\rho_v$</td>
<td>0.35</td>
</tr>
<tr>
<td>Voice BIR</td>
<td>$\gamma_v$</td>
<td>4 dB</td>
</tr>
<tr>
<td>Average talk spur</td>
<td>$T_c$</td>
<td>0.7 s</td>
</tr>
<tr>
<td>Data rate (I)</td>
<td>$R_d.1$</td>
<td>14.4 Kbps</td>
</tr>
<tr>
<td>Data delay deadline (I)</td>
<td>$d_1$</td>
<td>100 ms</td>
</tr>
<tr>
<td>Data arrival rate per frame (I)</td>
<td>$\lambda_d.1$</td>
<td>144 bits</td>
</tr>
<tr>
<td>Data rate (II)</td>
<td>$R_d.2$</td>
<td>144 Kbps</td>
</tr>
<tr>
<td>Data delay deadline (II)</td>
<td>$d_2$</td>
<td>50 ms</td>
</tr>
<tr>
<td>Data arrival rate per frame (II)</td>
<td>$\lambda_d.2$</td>
<td>1.44 Kbits</td>
</tr>
</tbody>
</table>
In our simulation, we assume that at each frame time the voice traffic has the highest priority and is always transmitted. The remaining resource (virtual bandwidth) after considering the voice is allocated among the two types of data users using the packet scheduling algorithm in section III. The assigned transmission rate (for flows with positive priority) has a lower limit of $R_{min} = 14.4$ Kbps. For simplicity we assign each data rate as multiples of the “basic” rate 14.4 Kbps.

Setting the outage probability $\delta = 0.1$, from (7) we know that $N_{v} = 43$ voice users, $N_{d,1} = 3$ type I data users, and $N_{d,2} = 2$ type II data users can be accommodated in the cell. Operating on these parameters, Table II shows the simulation results of our packet scheduling algorithm (the “with channel status” part), where we can see that the average rates of data users are very close to their target values (14.4 Kbps with type I data and 144 Kbps with type II data). The average power listed in this table is actually the normalized unit-rate transmission power $s_{j}/s$ defined in section III-B. To avoid the ambiguity of packet size in the measurement, the delay and loss rate are calculated with regard to “bit” instead of packets.

<table>
<thead>
<tr>
<th>User id</th>
<th>Type</th>
<th>Avg data rate (Kbps)</th>
<th>Data loss rate</th>
<th>Avg power</th>
<th>Avg data rate (Kbps)</th>
<th>Data loss rate</th>
<th>Avg power</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>I</td>
<td>14.35</td>
<td>0.0005</td>
<td>99.2</td>
<td>14.36</td>
<td>0.0029</td>
<td>184.8</td>
</tr>
<tr>
<td>2</td>
<td>I</td>
<td>14.49</td>
<td>0.0008</td>
<td>85.6</td>
<td>14.34</td>
<td>0.0007</td>
<td>179.1</td>
</tr>
<tr>
<td>3</td>
<td>I</td>
<td>14.42</td>
<td>0.0007</td>
<td>89.9</td>
<td>14.31</td>
<td>0</td>
<td>179.3</td>
</tr>
<tr>
<td>4</td>
<td>II</td>
<td>142.3</td>
<td>0.0010</td>
<td>92.4</td>
<td>143.9</td>
<td>0</td>
<td>179.3</td>
</tr>
<tr>
<td>5</td>
<td>II</td>
<td>143.8</td>
<td>0.0010</td>
<td>103.7</td>
<td>144.8</td>
<td>0</td>
<td>203.2</td>
</tr>
</tbody>
</table>

In future work we will consider the effects of on other-cell interference and imperfect power control in the packet scheduling.

REFERENCES