Multiuser Scheduling for High Speed Uplink Packet Access

Raymond Kwan†, Cyril Leung††, and Jie Zhang†
†Centre for Wireless Network Designs (CWIND), University of Bedfordshire, Luton, UK
††Dept. of Electrical & Computer Engineering, The University of British Columbia, B.C., Canada
Emails: raymond.kwan@beds.ac.uk, cleung@ece.ubc.ca, Jie.zhang@beds.ac.uk

Abstract—The problem of efficiently allocating resources to High Speed Uplink Packet Access (HSUPA) users in a 3GPP network is studied. A discrete optimization problem is formulated in which user channel qualities and buffer sizes are used at the base station to jointly allocate uplink resources. To circumvent the possibility of local optima, a linearized version of this formulation is also presented. It is shown that knowledge of user buffer sizes can significantly improve the overall achievable bit rate.

I. INTRODUCTION

Due to the ubiquity of Internet access and technological advances in mobile communications, wireless multimedia applications have become a reality. However, such applications can be very resource demanding. Due to the nature of the applications, many types of traffic can be encountered with different characteristics and Quality of Service (QoS) requirements [1]. The ability to efficiently allocate limited radio resources to the different traffic types while meeting their QoS requirements (resource scheduling) is a difficult task.

Since the traffic is expected to be dominated by web-browsing and file downloads, current wireless network design places a heavier emphasis on the downlink rather than the uplink. For example, the High Speed Downlink Packet Access (HSDPA) protocol was introduced in the Third Generation Partnership Project (3GPP) standard in order to provide higher bit rates on the downlink [2]. However, the increasing popularity of applications involving file uploads suggests that higher bit rates also need to be supported on the uplink. Studies on uplink scheduling have been reported in [3]-[8].

Recently, the High Speed Uplink Packet Access (HSUPA) protocol has been proposed in 3GPP in order to improve the supported channel bit rate on the uplink. It is essentially an enhanced version of the traditional dedicated channel (DCH) in the Wideband Code-Division Multiple Access (WCDMA) standard, and is a counterpart to the more established HSDPA technology. In contrast to HSDPA, HSUPA is not based on the shared-channel concept, and supports soft and softer handover [2]. In addition, the HSUPA scheduling functionality is moved from the Radio Network Controller (RNC) to the base station (Node-B in WCDMA standard terminology) in order to reduce signalling latency. Research efforts related to HSUPA can be found in [9]-[11] and the references therein.

In this paper, a novel joint multiuser optimization formulation for HSUPA is proposed which takes into account not only the channel qualities of the users, but also the amount of data in their buffers (buffer sizes). Such an approach allows the base station to make better uplink scheduling decisions by allocating resources to users more efficiently.

The remainder of the paper is organized as follows. In Section II, the system model is discussed, followed by the proposed optimization formulations in Sections III and IV. Numerical results are provided in Section V and the main findings are summarized in Section VI.

II. SYSTEM MODEL

Unlike HSDPA, where the power and Orthogonal Variable Spreading Factor (OVSF) code resources are centralized at the base station and are allocated to a few selected users, the uplink signal powers due to the mobiles are more evenly distributed, and the total received power is often viewed as the network resource in the uplink [1].

Let \( N \) be the total number of HSUPA users. The received Signal-to-Interference plus Noise Ratio (SINR) for an HSUPA user \( i \in \{1, \ldots, N\} \) is given by

\[
\gamma_i = \frac{g_i p_i}{\sum_{j \neq i} g_j p_j + I_T}, \quad \forall i,
\]

where \( p_i \) is the Enhanced Dedicated Physical Data CHannel (E-DPDCH)\(^1\) transmit power for user \( i \) and \( g_i \) is the path gain between the base station and user \( i \). The term \( I_T \) is the total received interference and noise power excluding the power due to E-DPDCHs within the target cell. After some algebraic manipulations, the transmit power for user \( i \) can be expressed as

\[
p_i = \frac{a_i / g_i}{1 - \sum_{i=1}^{N} a_i} I_T, \quad \forall i,
\]

where the term \( a_i \) is known as the load factor [1], and can be expressed as

\[
a_i = \frac{\gamma_i}{1 + \gamma_i}, \quad \forall i.
\]

It can be seen from (2) that the total load factor must satisfy the condition

\[
\sum_{i=1}^{N} a_i < 1.
\]

\(^1\)The term E-DPDCH is used to refer to the physical-layer channel in HSUPA on which user data are sent.
III. PROBLEM FORMULATION

Let $\tilde{R}_i = \frac{D_i}{T}$ be the requested bit rate, where $D_i$ is the amount of data in the buffer of user $i$, and $T$ is the scheduling period. In order to provide higher peak rates and more bit rate selections, multiple OVSF codes, or multicodes, can be used [2]. Let $\{G_{i,m_i}, m_i = 1, \ldots, M_{i,max}\}$ be $M_{i,max}$ possible spreading factors associated with the multicodes for user $i$, and $K_{i,max}$ be the maximum number of multicodes for user $i$. Thus, the complete set of bit rates that user $i$ can have is given by $\{R_i = k_i W / G_{i,m_i}, m_i \in M_i, k_i \in K_i\}$, where $W$ is the chip rate, $M_i = \{1, \ldots, M_{i,max}\}$ and $K_i = \{1, \ldots, K_{i,max}\}$. Depending on different mobile capabilities, only a subset of these bit rates are typically allowed [2]. Let $J_i$ be the number of possible bit rates for user $i$ chosen from the allowed combinations of spreading factors $M_i \subseteq M_i$ and the number of multicodes $K_i \subseteq K_i$ [12]. In other words, $J_i = |M_i| |K_i| \leq |M_i| |K_i|$. In this paper, the uplink power is assumed to be distributed equally among all multicodes for each user. Furthermore, let

$$G_{i,j} = \frac{G_{i,m_j}}{k^*_i}, \quad j = 1, \ldots, J_i, \quad i = 1, \ldots, N \quad (5)$$

be the fractional spreading factors, where $m^*_i \in M_i$, and $k^*_i \in K_i$, and the received bit energy to interference and noise power spectral density per code for user $i$ be $\rho_i = G_{i,j} a_i$ for a chosen value of $j$. Finally, let $\rho^*_i$ be the required value of $\rho_i$ in order to satisfy certain mobile and bearer-specific QoS-related requirements, and is assumed to be known at the base station. The goal is to maximize the objective function

$$\max_B \left\{ \sum_{i=1}^N \xi_i \min \left( \sum_{j=1}^{J_i} \frac{b_{i,j} W}{G_{i,j}}, \tilde{R}_i \right) - \sum_{i=1}^N \xi_i a_i \right\} \quad (6)$$

subject to

$$a_i = \frac{\rho^*_i}{\rho^*_i + \sum_{j=1}^{J_i} b_{i,j} G_{i,j}}, \quad \forall i \quad (7)$$

$$\sum_{j=1}^{J_i} b_{i,j} = 1, \quad \forall i \quad (8)$$

$$0 \leq a_i \leq \left( 1 - \sum_{i=1}^N a_i \right) Q_i, \quad \forall i \quad (9)$$

$$\sum_{i=1}^N a_i \leq \Lambda, \quad \forall i \quad (10)$$

$$b_{i,j} \in \{0, 1\}, \quad \forall i, j \quad (11)$$

where $B = \{b_{i,j}\}$, $Q_i = g_i \tilde{P}_i / I_T$, and $\tilde{P}_i$ is the maximum total allowed E-DPDCH transmit power for user $i$. The first constraint (7) can be obtained directly from (3), together with (5) and the definition of $\rho_i$, while the constraint in (9) comes from the fact that $p_i \leq \tilde{R}_i$, with $p_i$ defined in (2). The term $\Lambda < 1$ in (10) represents the maximum allowable value of the total load factor, which are determined by radio resource management algorithms at the higher layers.

The first term in (6) corresponds to the weighted aggregate bit rates for all users, and the second term corresponds to the weighted aggregate load factor. The sets $\{\xi_i, \xi_i, i = 1, \ldots, N\}$ describe the relative degree of importance between the bit rates and load factors among the users within the objective function. For example, a small value of $\xi_i / \xi_i$ places a stronger emphasis on maximizing the bit rate over the load factor for user $i$ and vice versa. The actual values of these parameters can be set by the upper layers in order to prioritize users according to their Quality of Service (QoS) requirements.

The goal of the above optimization problem is to obtain the optimal solution to (6) by selecting an appropriate matrix $B$ subject to constraints (7)-(11). Once $B$ is known, the load factors $a_i, i = 1, \ldots, N$ can be computed using (7). Since the total load factor $\sum_{i=1}^N a_i$ is related to the total received noise rise at the base station, the term $a_i$ represents the optimal share of noise resource that user $i$ is allocated. Subsequently, the allocated powers, $\{p_i, i = 1, \ldots, N\}$, can be obtained using (2) and communicated to the users, for example, via the Enhanced Absolute Grant Channels (E-AGCH) [13].

Note that the total received non E-DPDCH power, $I_T$, the path gain, $g_i$, and the requested bit rates, $\tilde{R}_i$, are assumed to be known or estimated at the base station. One possible way to estimate $g_i$ is based on the UE transmission Power Headroom (UPH), $\eta_i$, within the scheduling information (SI) portion of the MAC-e Packet Data Unit (PDU) as described in [13]. The UPH is defined as the ratio of the maximum possible uplink transmit power, $P^{max}_i$, and the transmit DPCCH code power, $P^{max}_i$, i.e. the power associated with a single OVSF code channel [14]. In other words, $g_i$ can be obtained as $g_i = \eta_i P^{max}_i / P^{max}_i$. Once the received DPCCH code power, $P^{max}_i$, is known at the base station. The requested bit rates, $\tilde{R}_i$, can be obtained or estimated based on the Total E-DCH Buffer Status (TEBS) bits and/or the Highest priority Logical channel Buffer Status (HLBS) bits within the SI portion of the MAC-e PDU [13].

IV. LINEARIZED FORMULATION

The optimization problem in (6) is an Integer Non-linear Programming (INP) problem. Besides the high computational complexity, one problem with solving such a formulation using standard methods such as the Branch-and-Bound is the possibility of obtaining locally optimal solutions. This problem can be avoided by re-formulating the problem as an Integer Linear

2Without loss of generality, we assume that $\tilde{R}_i$ corresponds to the physical layer bit rate. The relationship between $\tilde{R}_i$ and the requested bit rate at the Medium Access Control (MAC) level is assumed to be known, and is not considered in this paper.
Programming (ILP) problem as follows:

$$\max_B \sum_{i=1}^{N} (\zeta_i t_i - \xi_i a_i)$$

subject to

$$t_i \leq \sum_{j=1}^{J_i} b_{i,j} W_{i,j}, \quad \forall i$$ \hspace{1cm} (13)

$$t_i \leq \bar{R}_i, \quad \forall i$$ \hspace{1cm} (14)

$$\sum_{j=1}^{J_i} b_{i,j} = 1, \quad \forall i$$ \hspace{1cm} (15)

$$0 \leq a_i \leq \left( 1 - \sum_{i=1}^{N} a_i \right) Q_i, \quad \forall i$$ \hspace{1cm} (16)

$$\sum_{i=1}^{N} a_i \leq \Lambda,$$ \hspace{1cm} (17)

$$\rho_i^* = \rho_i^* a_i + \sum_{j=1}^{J_i} m_{i,j} G_{i,j}, \quad \forall i$$ \hspace{1cm} (18)

$$m_{i,j} \leq a_i + M (1 - b_{i,j}), \quad \forall i, j$$ \hspace{1cm} (19)

$$m_{i,j} \geq a_i - M (1 - b_{i,j}), \quad \forall i, j$$ \hspace{1cm} (20)

$$m_{i,j} \leq M b_{i,j}, \quad \forall i, j$$ \hspace{1cm} (21)

$$m_{i,j} \geq -M b_{i,j}, \quad \forall i, j$$ \hspace{1cm} (22)

$$b_{i,j} \in \{0, 1\}, \quad \forall i, j$$ \hspace{1cm} (23)

Note that the product $a_i b_{i,j}$, which results if both sides of (7) are multiplied by $\rho_i^* + \sum_{j=1}^{J_i} b_{i,j} G_{i,j}$, can be eliminated by introducing the variable $m_{i,j} = a_i b_{i,j}$, which is modeled by constraints (19)-(22), where $M$ is an arbitrarily large real value. Also, the minimum function in (6) can be linearized by introducing the variable $t_i$, together with constraints (13)-(14).

The above technique of introducing auxiliary variables in order to linearize a discrete optimization problem is commonly used in operations research [15], and is not discussed in detail here due to space limitation.

\[ \text{V. Numerical Results} \]

Let $\tilde{R} = [\tilde{R}_1, \tilde{R}_2, \ldots, \tilde{R}_N]$, $\rho = [\rho_1, \rho_2, \ldots, \rho_N]$, and $Q = [Q_1, Q_2, \ldots, Q_N]$. Furthermore, let the total normalized achievable bit rate $R_{tot}$, the total normalized allocated bit rate $R_{alloc}$, and the total load factor $\Lambda_{tot}$ be

$$R_{tot} = \frac{1}{W} \sum_{i=1}^{N} \min \left( \sum_{j=1}^{J_i} b_{i,j} W_{i,j}, \bar{R}_i \right)$$ \hspace{1cm} (24)

$$R_{alloc} = \frac{1}{W} \sum_{i=1}^{N} \sum_{j=1}^{J_i} b_{i,j} W_{i,j} G_{i,j}$$ \hspace{1cm} (25)

$$\Lambda_{tot} = \sum_{i=1}^{N} a_i$$ \hspace{1cm} (26)

Ideally, $R_{tot}$ should be as close to $R_{alloc}$ as possible. For simplicity, let the average E-DPDCH channel quality $\bar{Q} = \frac{\sum_{i=1}^{N} Q_i}{N}$ represent the aggregate effect of the maximum uplink transmit power, the path gain between the base station and the user, and the non-E-DPDCH interference received at the base station. A higher value of $\bar{Q}$ implies a more favorable condition for HSUPA users, and vice versa.

To illustrate the importance of traffic load aware scheduling, the above load-and-channel based (LCB) scheme is compared with a channel-based (CB) scheme in which only the instantaneous channel qualities are taken into considerations, i.e. the scheduling decision is independent of the requested bit rates. Then, the objective in (6) reduces to

$$\max_B \left\{ \sum_{i=1}^{N} \left( \zeta_i \sum_{j=1}^{J_i} b_{i,j} W_{i,j} - \xi_i a_i \right) \right\}.$$ \hspace{1cm} (27)

As a simple illustration, let $N = 3$, and $G_i = [G_{i,1}, G_{i,2}, \ldots, G_{i,9}] = [2, 4, 8, 16, 32, 64, 128, 256, 512]$, $\forall i$. The objective function in (27) reduces to $R_{alloc}$ if $\zeta_i = 1$ and $\xi_i = 0$ for all $i$.

Figures 1 and 2 show the total achievable bit rate, $R_{tot}$, and the total load factor, $\Lambda_{tot}$, as a function of the average E-DPDCH channel quality, $\bar{Q}$, for the LCB scheduler. In these figures, it is assumed that $Q_i = [1, 0.5, 0.25]$ and $\rho = [5, 5, 5]$. Two cases are considered: Case 1 and Case 2 corresponding to $\bar{R} = [0.3, 0.15, 0.075]$ and $\bar{R} = [0.075, 0.15, 0.3]$ respectively.

In the first case, the user with the best channel quality also has the most data in the buffer, and vice versa. In the second case, the opposite is true. As expected, the total bit rate $R_{tot}$ increases as $\bar{Q}$ improves. More interestingly, it can be seen in Fig. 1 that the requested bit rate vector $\bar{R}$ can have a large impact on the performance. It is generally more resource efficient to allocate higher bit rates to users with better channel qualities. Thus, as illustrated in Fig. 1, a system in which users with good channel qualities have more data to send is more resource efficient than the reverse situation. As expected, $R_{tot}$ decreases when the load factor limit, $\Lambda$, is reduced. Comparing Figs. 1 and 2, it can be seen that no further increase in $R_{tot}$ occurs once $\Lambda_{tot}$ approaches $\Lambda$, even at a favorable channel quality. Note that $\Lambda_{tot}$ is not monotonically increasing with $\bar{Q}$ since the objective is to maximize $R_{tot}$ and not $\Lambda_{tot}$.

Figs. 3 and 4 show $R_{tot}$ and $\Lambda_{tot}$ respectively as a function of the average E-DPDCH channel quality, $\bar{Q}$, for the LCB and the CB schedulers under the case 1 scenario. Recall that this scenario corresponds to the situation where users with better channel qualities have more data to send. Intuitively, one might feel that in this case, the CB scheduler would have a similar $R_{tot}$ as the LCB scheduler. However, as can be observed in Fig 3, this is not always the case. It is possible for the CB scheduler to achieve an optimal value of $R_{alloc}$, but a sub-optimal value of $R_{tot}$. The $R_{tot}$ achieved by the CB scheduler does not always increase monotonically with $\bar{Q}$. As shown in Fig. 4, the $\Lambda_{tot}$ achieved by both schedulers are similar.
VI. Conclusion

In this paper, a joint multiuser optimization formulation has been proposed, in which the channel qualities and the amount of data in the user buffer, $D_i$, of the users are used at the base station to jointly allocate uplink resources for the users. In order to avoid the possibility of local optima, a linearized version of the formulation was also presented. Numerical results show that knowledge of $D_i$ can be used to significantly improve the overall achievable bit rate.

Acknowledgment: This work was supported in part by a Marie Curie Post-Doctoral Fellowship, the Natural Sciences and Engineering Research Council (NSERC) of Canada under Grant OGP0001731, and by the UBC PMC-Sierra Professorship in Networking and Communications.

REFERENCES


Fig. 1. Total normalized allocated bit rate, $R_{\text{tot}}$, as a function of the average E-DPDCH channel quality, $\bar{Q}$, for the LCB scheduler. Case 1: $\bar{R} = [0.3, 0.15, 0.075]$; Case 2: $\bar{R} = [0.075, 0.15, 0.3]$.

Fig. 2. Total load factor, $\Lambda_{\text{tot}}$, as a function of the average E-DPDCH channel quality, $\bar{Q}$, for the LCB scheduler. Case 1: $\bar{R} = [0.3, 0.15, 0.075]$; Case 2: $\bar{R} = [0.075, 0.15, 0.3]$.

Figs. 5 and 6 show $R_{\text{tot}}$ and $\Lambda_{\text{tot}}$ respectively as a function of the average E-DPDCH channel quality, $\bar{Q}$, for the LCB and CB schedulers under the Case 2 scenario. A comparison of Fig. 3 and Fig. 5 show that the difference between the $R_{\text{tot}}$ values for the LCB and CB schedulers is much higher for case 2 than for case 1. This is because under the Case 2 scenario, the CB scheduler attempts to allocate lower spreading factors or higher power to users with good channel qualities, even though these users have less data to send.
Fig. 3. Total normalized allocated bit rate, $R_{\text{tot}}$, as a function of the average E-DPDCH channel quality, $Q$, for the LCB and CB schedulers, with $\tilde{R} = [0.3, 0.15, 0.075]$.

Fig. 4. Total load factor, $\Lambda_{\text{tot}}$, as a function of the average E-DPDCH channel quality, $Q$, for the LCB and CB schedulers, with $\tilde{R} = [0.3, 0.15, 0.075]$.

Fig. 5. Total normalized allocated bit rate, $R_{\text{tot}}$, as a function of the average E-DPDCH channel quality, $Q$, for the LCB and CB schedulers, with $\tilde{R} = [0.075, 0.15, 0.3]$.

Fig. 6. Total load factor, $\Lambda_{\text{tot}}$, as a function of the average E-DPDCH channel quality, $Q$, for the LCB and CB schedulers, with $\tilde{R} = [0.075, 0.15, 0.3]$. 