Normalized Rate Guarantee Scheduler for High Speed Downlink Packet Access

Kamal Deep Singh
IRISA/INRIA
Campus de Beaulieu
35042 Rennes Cedex, France
ksingh@irisa.fr

David Ros
GET/ENST Bretagne
Rue de la Châteaigneraie, CS 17607
35512 Cesson Sévigné Cedex, France
david.ros@enst-bretagne.fr

Abstract—High Speed Downlink Packet Access (HSDPA) is a packet-based data service in UMTS networks that supports data rates of several Mbit/s, making it suitable for applications ranging from file transfer to multimedia streaming. One of the salient points of HSDPA is the use of MAC-layer scheduling to perform resource management (i.e., bandwidth allocation between terminals), taking into account the radio channel conditions of all users. Usually, additional factors like fairness between users, cell throughput or quality-of-service (QoS) parameters are also considered in the scheduling mechanism.

In this paper we consider different QoS classes or user groups and study resource allocation among those user groups using HSDPA MAC-layer scheduling. We propose a Normalized Rate Guarantee (NRG) scheduler, which is an extension of a previous QoS scheduler by Hosein (2002). When allocating the available bandwidth, the NRG scheduler takes into account the fact that QoS requirements (in terms of minimum bandwidth) may vary from one flow to the other. It tries to apportion loss rates in a fairer way during congestion. Another goal of NRG is to avoid the deterioration of QoS when best-effort load is increased.

I. INTRODUCTION

Universal Mobile Telecommunications System (UMTS) cellular networks offer higher data rates with respect to older 2G and 2.5G networks and, with the Release 5 version, is evolving into an all-IP packet network. Moreover, recent enhancements have made it possible for UMTS networks to support data rates of several Mbit/s, making them suitable for applications ranging from file transfer to multimedia streaming. A typical UMTS network consists of a core network and the UMTS Terrestrial Radio Access Network (UTRAN). The UTRAN consists of Radio Network Controllers (RNC) each controlling several base stations (BS, or Node B). A mobile user connects her User Equipment (UE) to the UTRAN, which in turn is connected to the Internet through the Serving GPRS Support Node (SGSN) and the Gateway GPRS Support Node (GGSN) present in the core network. In general, the links in the core UMTS network are over-provisioned. Such over-provisioning of the core network, together with the fluctuations in radio channel quality that are inherent of wireless links, will often make the UTRAN act as a bottleneck. Therefore, resource management will be required in UTRAN to provide good Quality of Service (QoS) to the users.

High Speed Downlink Packet Access (HSDPA) [1] is an enhancement to UMTS networks and it supports data rates of several Mbit/s. In spite of the fairly high data rates that HSDPA offers, the shared downlink radio channel used in HSDPA is a challenging environment for effective data scheduling by taking into account the different QoS requirements.

In this paper we propose a new scheduler called Normalized Rate Guarantee (NRG). NRG improves upon the existing HSDPA QoS schedulers because, unlike those schedulers, it is not sensitive to the best-effort (BE) load and during congestion it apportions the loss rates, among different QoS users, in a fairer way. The performance of NRG is evaluated using traces of real video encoded using the H.264 format. The remainder of this paper is organized as follows. Section II discusses the related work. Some key concepts of HSDPA and some details of our simulation platform are given in sections III and IV. We describe the NRG scheduler in section V and its performance is evaluated in section VI. Section VII concludes the paper.

II. RELATED WORK

Basic channel-adaptive Proportionally Fair (PF) scheduling that allocates resources among all users is discussed by Jalali et al. in [2]. Channel-adaptive schedulers are reported to take advantage of the multi-user diversity that is inherent of wireless systems. Borst [3] studied the user-level performance of the PF scheduler and gave analytical results by considering scheduling strategies for the QoS users. A dynamic setting with a random nature of service demands is considered in [3], in comparison to many previous works assuming static user population. Bonald et al. [4] extended the results of [3] by using some general scheduling and admission control schemes. Andrews et al. proposed a scheduler based on providing delay and rate guarantees in [5]. Barriac et al. [6] extended the PF scheduler to equalize the user throughputs for QoS users. Hosein showed in [7] that “good” schedulers can be designed based upon user utility functions. Kolding [8] proposed a new scheduler, based on a better estimation of the required scheduling activity for a given user. Pedersen et al. [9] discussed the mapping of the schedulers to the QoS parameters specified by the 3GPP project.

III. HIGH SPEED DOWNLINK PACKET ACCESS (HSDPA)

A new transport channel called High Speed Downlink Shared Channel (HS-DSCH) has been introduced for HSDPA...
HS-DSCH is supported by an auxiliary channel called High Speed Shared Control Channel (HS-SCCH). The goal of the latter is to allow for fast monitoring of the radio channel conditions of all users: every 2 ms, a UE can send to the base station a Channel Quality Indicator (CQI) over this control channel. This feedback makes it possible to optimize the transmission by means of channel-adaptive schemes. Thus, the CQI is used to adapt the coding rate, modulation scheme and number of codes employed, so that users having good channel conditions may be provided with high data rates. In UMTS release’99, frames with errors are retransmitted by the RLC (Radio Link Control) protocol layer present in the RNC. In HSDPA, the MAC layer itself can perform fast retransmission of the erroneous data, and retransmission times are reduced as this functionality has been moved to the BS. This also provides fast response times of the channel-adaptive schemes described above. At the MAC level, fast Hybrid-ARQ (HARQ) is used to retransmit the erroneous frames; UEs in turn do not discard the erroneous frames, but combine them with the successive retransmitted frames using schemes like Chase Combining and Incremental Redundancy.

HSDPA is a shared wireless channel and providing resource management (that is, allocating bandwidth among terminals) is complex due to variable radio channel conditions; these may change rapidly, especially for users farther from the BS. Resource management is thus performed by a scheduler that may adapt to fluctuating radio conditions. The Transmission Time Interval (TTI) in HSDPA has been reduced to 2 ms as compared to the 10, 20, 40, and 80 ms intervals supported by UMTS Release’99. Every TTI, the scheduler chooses the next user to be served based on the channel conditions of all the users. This channel-adaptive scheduling can be used, for instance, to maximize the global cell throughput by scheduling users only when their channel condition is good. Note that a given UE can experience strong fluctuations in bandwidth over time, due to such channel-dependent scheduling policies.

A. Simulation Platform

We simulate HSDPA using the well-known ns-2 simulator [10], compiled with the EURANE extensions [11]. EURANE models the UTRAN in detail, in particular, it implements all the functions of the RLC and MAC layers.

It should be remarked that there are per-user queues in the RNC, so the degree of statistical multiplexing in those queues is very low. The simulated MAC layer implements the HSDPA scheduler and other functionalities like HARQ. The underlying physical layer is modeled in detail, as described in [12], and this model is used to compute a CQI value which is fed back from UEs to the base station. The default values for some of the configuration parameters can be found in the Table I.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Default Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cell Radius</td>
<td>500 m</td>
</tr>
<tr>
<td>Multipath fading environment</td>
<td>Pedestrian A, 3.0 km/h</td>
</tr>
<tr>
<td>$\tau$ in Equation (1)</td>
<td>1000</td>
</tr>
<tr>
<td>RLC mode</td>
<td>Unacknowledged (UM)</td>
</tr>
<tr>
<td>RTT Internet</td>
<td>60 to 100 ms (variable)</td>
</tr>
<tr>
<td>RTT UMTS core</td>
<td>50 ms</td>
</tr>
<tr>
<td>Single Simulation Run</td>
<td>72 seconds</td>
</tr>
<tr>
<td>BE flows (Long-lived, TCP flows)</td>
<td>Newreno</td>
</tr>
<tr>
<td>Packet Size, Receiver Window</td>
<td>1000 bytes, 10000 pks</td>
</tr>
<tr>
<td>RIO parameters [13]</td>
<td></td>
</tr>
<tr>
<td>$\min_{thred,\minthyellow,\mindhgreen}$</td>
<td>(0.1, 0.3, 0.5) * $Q$</td>
</tr>
<tr>
<td>$\max_{thred,\maxthyellow,\maxdhgreen}$</td>
<td>(0.3, 0.5, 0.7) * $Q$</td>
</tr>
<tr>
<td>$\maxP_{red,\maxP_{yellow,\maxP_{green}}}$</td>
<td>0.2, 0.1, 0.02</td>
</tr>
<tr>
<td>$w_p$</td>
<td>0.9</td>
</tr>
</tbody>
</table>

IV. SCHEDULING IN HSDPA

The HSDPA scheduler is the key to resource management in the UTRAN downlink, because it decides which user is to be scheduled at each time slot. A simple round-robin (RR) scheduling gives the time slot to the users in a round-robin manner, i.e., is fair with respect to system resources (time slots). However, this policy is not optimal as it does not take into account the user channel conditions. Maximum C/I (CI) scheduling gives the channel to the user having the best channel conditions at each given time slot. If $R_i(t)$ is the instantaneous data rate experienced by user $i$ at time $t$, then the CI scheduler assigns the slot at time $t$ to a user $i^*$ such that: 

$$i^* = \arg \max_i \{ R_i(t) \},$$

that is, it gives the channel to the user able to achieve the highest instantaneous rate. The CI scheduler provides the highest cell (global) throughput but is very unfair as a user closest to the base station can get all the resources, and the users farther away will be starved. The Proportionally Fair (PF) scheduling algorithm [2], [14] assigns the slot at time $t$ to a user $i^*$ such that:

$$i^* = \arg \max_i \{ R_i(t)/\lambda_i(t) \},$$

with $\tau > 1$ and $\Delta t$ equal to the length of the TTI. The PF scheduler offers a good trade-off between cell throughput and fairness, as it both gives the channel to the user having “relatively good” channel conditions and provides the so-called “proportional fairness” defined in [15].

QoS schedulers in general pick a user $i^*$ satisfying:

$$i^* = \arg \max_i \{ B_i(t) R_i(t)/\lambda_i(t) \}$$

where $B_i(t)$ represents a “barrier function” [7]. In order to design QoS schedulers, it was shown in [7] that, given user utilities $U_i(\lambda_i)$, it is possible to obtain a “good” scheduler that will pick the user $i^*$ such that:

$$i^* = \arg \max_i \{ R_i(t) \cdot U_i'(\lambda_i(t)) \}$$

Note that, in the above formulation, the corresponding utility functions for RR, CI and PF scheduling are $U(\lambda) = 1$, $U(\lambda) = \lambda$ and $U(\lambda) = \log \lambda$, respectively.
A Rate-Guarantee (RG) scheduler was thus designed by Hosein in [7] using Eq. (3). A utility function

\[ U_Q(\lambda) = \log(\lambda) + 1 - \exp(-\beta \cdot (\lambda - \lambda_{\text{min}})) \quad (4) \]

was assumed for QoS users (with \( \beta > 0 \)), resulting in a scheduler with a barrier function

\[ B_i(t) = \begin{cases} 1 + \lambda_i(t) \beta \cdot \exp(-\beta \cdot (\lambda_i(t) - \lambda_{\text{min}}^{(i)})) & \forall i \in Q, \\ 1 & \forall i \in B, \end{cases} \]

which corresponds to the case where a minimum throughput \( \lambda_{\text{min}}^{(i)} \) is required by a QoS user \( i \); \( Q \) and \( B \) denote the set of QoS and best-effort (BE) users, respectively. Note that the \( B_i(t) \) function for BE users corresponds to a user utility of the form \( U_B(\lambda) = \log(\lambda) \).

In the literature, there exist several studies (e.g., [7], [9], [16]) that use an RG scheduler in order to provide rate guarantees to QoS users. In [16], a different variant of the RG scheduler described by (5) is used with the following change for QoS users:

\[ B_i(t) = 1 + \alpha \exp(-\beta \cdot (\lambda_i(t) - \lambda_{\text{min}}^{(i)})) \quad \forall i \in Q; \quad (6) \]

note that \( \alpha \) is a constant, independent of both \( \lambda_i(t) \) and \( \lambda_{\text{min}}^{(i)} \). The scheduler given by (6) is also used as a reference scheduler in [8].

V. NORMALIZED RATE GUARANTEE SCHEDULING

For any user \( i \in Q \), let us denote by \( \Delta \lambda_i(t) \) the instantaneous difference at time \( t \) between the average rate \( \lambda_i(t) \) and its minimum guaranteed rate \( \lambda_{\text{min}}^{(i)} \), that is, \( \Delta \lambda_i(t) = \lambda_i(t) - \lambda_{\text{min}}^{(i)} \). The purpose of the barrier functions (5)-(6) is then to increase the probability of scheduling user \( i \) whenever \( \Delta \lambda_i(t) < 0 \).

Both the original RG scheduler in [7] and its variant in [16] suffer from the deterioration of rate guarantees of the QoS users as the number of “active” BE users \( n_{BE} \) is increased. This is because, as \( n_{BE} \) increases, the value of \( \lambda_i \) for BE users decreases. This in turn increases the term \( B_i(t)/\lambda_i \) for BE users as \( B_i(t) = 1 \) remains constant. Moreover, the above schedulers, especially the one described by (6), tend to be biased towards some users depending on their value of rate guarantees. This is further discussed in Section VI.

We propose a new scheduler called Normalized Rate Guarantee (NRG) that is based on the RG scheduler. The rationale of NRG scheduler is that we want to apportion loss rates in a fairer way during congestion irrespective of the rate guarantees. We also want that the increase in the BE load does not deteriorate the quality of the QoS flows and the increased BE load be shared among BE users only.

We assume the following utility for QoS users:

\[ U_Q(\lambda) = \lambda_{\text{min}} \cdot \left( \log(\lambda) + 1 - \exp\left(-\beta \frac{\lambda - \lambda_{\text{min}}}{\lambda_{\text{min}}}\right) \right). \quad (7) \]

For BE users we consider that rate guarantees are always “satisfied”, so we assume a utility for BE users of the form:

\[ U_B(\lambda) = \frac{k_{BE} n_{BE}}{n_{BE}} \log(\lambda). \quad (8) \]

The constant \( k_{BE} \) will determine the proportion of resources allocated to the BE users; the \( n_{BE} \) term, similar to the concept used in [8], will ensure that if the number of BE users increases, then it will not increase the load on resources for QoS users. Furthermore, a value \( n_{BE}^{(min)} \) can be chosen such that \( B_i(t) = k_{BE}/n_{BE}^{(min)} \) if \( n_{BE} < n_{BE}^{(min)} \) to avoid the division by zero and to slightly improve the QoS when the number of BE users goes low.

Using the above utilities (7)-(8) and Eq. (3) results in a scheduler of the form given by (2) with:

\[ B_i(t) = \begin{cases} \lambda_{\text{min}}^{(i)} + \lambda_i(t) \beta \exp(-\beta(\lambda_i(t) - \lambda_{\text{min}}^{(i)})/\lambda_{\text{min}}^{(i)}) & \forall i \in Q, \\ k_{BE}/n_{BE} & \forall i \in B. \end{cases} \]

The parameter \( \beta \) will determine the aggressiveness of the scheduler to the increasing losses (due to decreasing average rates). This is further discussed in section VI.

It should be noted that the term \( B_i(t)/\lambda_i(t) \) in (2) becomes equal for all the QoS users when either all users are just achieving their minimum throughput (i.e., \( \Delta \lambda_i = 0 \ \forall i \)), or when the loss ratio of QoS users is the same. Thus, we call it as normalized rate guarantee scheduler because both the terms \( \Delta \lambda_i(t) \) and \( B_i(t) \) for BE users are normalized.

VI. PERFORMANCE EVALUATION

In this section we evaluate the performance of the NRG scheduler. For all the results, a total of 40 independent simulation runs are performed to obtain good confidence. The HSDPA simulation platform described before is used for the simulations and all the video users and the BE users are randomly located inside the cell.

For the first set of simulations we use real video traces and use a trace-based methodology similar to that of [17]. This methodology permits us to use real videos for simulations over ns-2. The video used is shown in Fig. 1 and is encoded by the H.264 codec used in [17]. We assume four video users who stream (using UDP) a pre-encoded video file from a server located in the Internet. A total number of \( n_{BE} \) background flows, representing BE users, are present and the traffic model used is TCP long-lived flows.

In a typical video streaming scenario, there is a playout buffer at the client and the playout is started after some delay while the buffer gets filled. A typical range of values for the playout buffer is 4 to 8 seconds. We assume then a video playout buffer of 8 seconds and any video packet delayed more than 8 s is discarded. The “average” bitrate for two video users is 128 kbps and for the other two video users is 360 kbps. The target average bitrates of the encoded video streams are controlled using the quantization parameter of the H.264 codec. Fig. 2 shows the bitrate versus time for the 360 kbps video (the bitrate of the 128 kbps stream, not shown for space reasons, is qualitatively similar).
The values of $\beta$ is similar to the one used in [7], [16] when the rates are provided rate guarantees even on slight channel deterioration. It can be seen in Fig. 4(a) that the average loss rate almost doubles with the RG scheduler as $n_{BE}$ is increased, whereas it remains stable with the NRG scheduler. Note that we compare the byte loss rate ($L_B = \frac{\text{bytes lost}}{\text{bytes sent}}$) instead of packet loss rate because of the variable packet size. A similar observation can be made for the average peak signal-to-noise ratio (PSNR) values of the received videos in Fig. 4(b), i.e., with the RG scheduler the average PSNR deteriorates when the BE load is increased. The reason for this behavior is that when the BE load increases, the RG scheduler cannot avoid giving more resources to the BE flows, as explained by the increase in total BE throughput in Fig. 4(c), at the cost of QoS/video users. On the other hand, this deterioration is not there with the NRG scheduler. NRG, thus, improves upon the RG scheduler as it remains insensitive to the BE load as is clear in the figures.

Another set of simulations were run with six video users and $n_{BE} = 28$ best-effort users. The video flows are modeled by CBR sources. At the client side, the packets that are delayed by more than 4 s are dropped. The values of $\lambda_{\text{min}}$ and sending rates for the six video users are 32, 64, 128, 256, 384 and 512 kbps. Per-user drop-tail queues holding up to 2 s of video data are used at the RNC and video packet size is 500 bytes.

In order to look at the bias towards certain rate guarantees, we compare NRG, “NRG-Hosein” (i.e., taking $B_i(t)$ for QoS users from (5) and for BE users from (9)) and “NRG-Lundevall” (i.e., taking $B_i(t)$ for QoS users from (6) and for BE users from (9)). For the last two “hybrid” schedulers, we took the $B_i$ from NRG, $\forall i \in B$, so as to remove the sensitivity to BE load, in order to focus only on the effect of $\lambda_{\text{min}}$ values. Nevertheless, the $B_i(t)$ functions for QoS users were not normalized for the last two schedulers. The value of $\beta$ for the last two schedulers is $3.125 \cdot 10^{-5}$ and $k_{BE}$ is 4.5 when the rates are in bps. The value of $\alpha$ is 1.25 for the “Lundevall” variant given by Eq. (6). The values of $\beta$ and $k_{BE}$ for NRG are the same as in the first set of simulations with video traces. For a fair comparison, the above parameters were chosen such that the loss rates were minimal and the capacity (total guaranteed rate + total TCP throughput) obtained was the same and equal to 2 Mbps. Fig. 5 shows the loss rates for different QoS
users. It can be seen that the NRG-Lundevall scheduler is biased towards the low bit rate users. NRG thus improves greatly upon NRG-Lundevall and slightly upon NRG-Hosein (as overall loss rate is less for NRG) in terms of providing fair loss rates to different QoS users.

VII. CONCLUSION

We proposed a new scheduler called Normalized Rate Guarantee (NRG) that is based on an existing Rate Guarantee (RG) scheduler. The performance evaluation was done using the traces of a real reference video encoded using the H.264 format. It was shown that the NRG scheduler improves upon the RG scheduler as it does not deteriorate the quality of the QoS flows when the load of the best-effort users is increased. Moreover, the NRG scheduler provides fair loss rates irrespective of the rate guarantees of the different QoS users.

As a next step we would like to test the NRG scheduler with more traffic scenarios. Moreover, we would like to study the streaming of scalable video (MPEG4-SVC) over UMTS/HSDPA networks.

ACKNOWLEDGMENT

This work was partially funded by the Conseil Régional de Bretagne, under grant B/1042/2004/MOTIPV6.

REFERENCES