

The Weighted Proportional Fair Scheduler

Kinda Khawam, Daniel Kofman
GET/ENST Telecom Paris
{khawam, kofman}@enst.fr

Eitan Altman
INRIA Sophia Antipolis
eitan.altman@sophia.inria.fr

Abstract

To this day, traffic in cellular networks is still mainly real-time. However, with the deployment of new technologies, such as HDR (High Data Rate) and HSDPA (High Speed Downlink Packet Access), this situation is bound to evolve rapidly and elastic traffic will significantly increase. These novel technologies implement opportunistic scheduling, taking advantage of the delay-tolerance of elastic traffic, to augment the global capacity of the system. Previous work has shown that "Proportional Fair" (PF) is an opportunistic scheduler that provides a good compromise between fairness and efficiency. Nevertheless, the hypotheses according to which these results are obtained are not always valid in real environments. In this paper, we propose a modified version of PF that not only introduces flexibility in sharing resources between active users but also allows for a fair allocation of resources in realistic environments.

1. Introduction

Data Traffic is increasingly popular in 3G mobile networks. New technologies like HDR [1] and its equivalent in 3GPP, HSDPA [6], offer higher data rates than previous architectures notably through opportunistic scheduling: time is divided into short intervals and the base station (BS) transmits at full power to a single user per time slot (intra-cell interference cancellation); the scheduler reaps the benefits of multi-user diversity over short time-scales and determines how the resources are allocated over longer time-scales.

A scheduler has usually two contradictory targets: maximizing the overall throughput and guaranteeing fairness. The well-know PF scheduler transmits to the user with the highest data rate relative to its present realized mean data rate in order to conjugate fairness and efficiency. Yet, PF is a rigid and non-adaptable scheduler as it falls short from enabling the system to define which trade-off

between efficiency and fairness is targeted. Moreover, in a homogeneous environment, PF has only a "fair power" sharing but can still be considered as "unfair" as the throughput perceived by users decreases with distance. More importantly, this restricted "fairness" is not fulfilled with heterogeneous fading. Hence, in a realistic environment, under PF scheduling, users with the most variable channel conditions receive the least amount of slots [4].

To cope with these drawbacks, we suggest in this paper an alternative scheduler, termed Weighted PF (WPF), which is a hierarchical scheduler that allows to fully control the trade-off between fairness and efficiency. To introduce the required control, we define different classes of users. At its first hierarchical level, WPF distributes the slots between the defined classes in a **Weighted** Round Robin (WRR) fashion. At its second level, users inside each class are served by means of PF (the PF scheduler takes independent decisions inside each class). We define the users' classes in such a way that, on the one hand, users belonging to a given class have comparable Signal-to-Noise Ratios (served according to PF, they will obtain comparable feasible rates) and, on the other hand, fading is homogeneous inside each class; thus, applying PF induces a strict fairness in accessing resources for users belonging to the same class (we stress on the fact that it is not possible to reach this target when applying PF to the whole cell due to heterogeneous fading). By doing so, not only will we control the resource allocation but also obtain in a real environment the behaviour that PF provides only in an idealistic environment.

The rest of the paper is organized as follows. In Section 2, we present the cell partitioning. In Section 3, we analyze the performances of the WPF scheduler; in particular, we obtain analytical results for the mean rate for a fixed number of users for PF and WPF in an environment with heterogeneous fading. In Section 4, we present an analytical study for WPF under dynamic traffic conditions. We conclude in Section 5.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

QShine '06 The Third International Conference on Quality of Service in Heterogeneous Wired/Wireless Networks
August 7-9 2006, Waterloo, ON, Canada
© 2006 ACM 1-59593-472-3/06/08...\$5.00

2 The cell Partitioning

To obtain different classes of users with comparable SNRs (Signal-to-Noise Ratio), we divide the cell into different geographical zones; each Zone z corresponding to the set of users whose distance to the BS ranges from a minimal value r_z to a maximal value r_{z+1} . We decided to segment the cell in only three zones ($z = 0, 1, 2$) since increasing the number of zones limits the number of users per zone and therefore reduces the gain resulting from multi-user diversity. Consequently, users in Zone 0 are those whose distance to the BS ranges between $r_0 = 0$ and r_1 , users in Zone 1 are those located between r_1 and r_2 , while users in Zone 2 are those located between r_2 and $r_3 = \mathcal{R}$, where \mathcal{R} is the ray of the cell. To justify our choice of the values taken by r_1 , r_2 and r_3 , we introduce the radio model we use.

2.1 The Radio Resource

In this section, we describe the model adopted for the radio resource and compute the feasible rate of each user accordingly.

2.1.1 The Propagation Model

The power received by a given user depends on the radio channel state and varies with time due to user mobility and fading effects (shadowing and multi-path reflections). In our model, the mobility will not be included, nor will the slow fading be.

Let P be the transmission power emitted by the BS, γ_k the free space path loss and x_k the fast fading (of unit mean) for user k . The power received by a user k situated at a distance r from the BS, at time t , is then given by:

$$P_k(r, t) = P \cdot \gamma_k(r) \cdot x_k(t) \quad (1)$$

The adopted model for the free space path loss is the following:

$$\gamma_k = 1 \text{ if } r \leq \epsilon \text{ and } \gamma_k = \left(\frac{\epsilon}{r}\right)^\beta \text{ otherwise}$$

where β is the path loss exponent (taking values between 2 and 5) and ϵ is the maximum distance at which the full power P is received.

2.1.2 The feasible rate

For user k , the Signal-to-Noise Ratio and Energy-per-bit to Noise Density Ratio [7] are respectively equal to:

$$SNR_k = \frac{P_k}{(\eta + I_k)}, \quad \frac{E_b}{N_0} = \frac{W}{R_k} \cdot SNR_k \quad (2)$$

where R_k is the feasible rate of user k , W the cell bandwidth, η the background noise and I_k the interference due to other BSs.

For a given target error probability, $\frac{E_b}{N_0}$ must be greater than a given threshold σ_k . Assuming the equality, the feasible data rate of a user k is then:

$$R_k = \frac{W}{\sigma_k} \cdot SNR_k$$

In the vast majority of references, σ_k is taken as a constant in order to preserve the linearity between R_k and SNR_k . This assumption is not valid when different types of modulation are used which is the case for HDR and HSDPA systems. Thus σ_k will vary with the feasible rate and hence with the distance from the BS. Therefore, we consider in our model different values of σ_k per zone (σ_k is then replaced by σ_z in R_k , for $z = 0, 1, 2$). In practice, the way we define the zones induces that inside Zone 0 and Zone 2, σ_k is indeed constant and equals 6.5dB and 2.5dB respectively [1]. For Zone 1, we define σ_1 as the mean value of σ_k in this zone. The feasible rate of user k in Zone z is then:

$$R_{k,z} = \frac{W}{\sigma_z} \cdot SNR_k$$

We denote by C_0 the maximum peak rate offered by the used coder and by r^* the maximum distance at which this peak rate is achieved in the absence of fading, i.e.,

$$r \leq r^* \Leftrightarrow R = C_0$$

We suppose that the interference I_k is constant inside each zone (it increases with the zone index). Hence, I_k will be replaced by I_z . Using (1), (2) and knowing that C_0 is the maximum peak rate that can be attained, we have the following:

$$R_{k,z}(r, t) = \min \left[C_0, \frac{W}{\sigma_z} \cdot \frac{P \cdot \gamma_k(r) \cdot x_k(t)}{(\eta + I_z)} \right] \quad (3)$$

Assuming that C_0 can be achieved, i.e. $r^* \geq \epsilon$, we get from the path loss model and (2) the following:

$$C_0 = \frac{W \cdot P}{\sigma_0 \cdot (\eta + I_0)} \cdot \left(\frac{\epsilon}{r^*}\right)^\beta \quad (4)$$

Using (3) and (4), the feasible rate of a user k in Zone z is therefore given by:

$$R_{k,z}(r, t) = C_0 \cdot \min \left[\left(\frac{r^*}{r}\right)^\beta \cdot x_k(t) \cdot K_z, 1 \right] \quad (5)$$

with $K_z = \frac{\sigma_0}{\sigma_z} \cdot \left(\frac{\eta + I_0}{\eta + I_z}\right)$.

The average feasible rate of a user k in Zone z , denoted by $C_{k,z}$, is then:

$$C_{k,z}(r) = C_0 \cdot \mathbb{E} \left[\min \left[\left(\frac{r^*}{r}\right)^\beta \cdot x_k(t) \cdot K_z, 1 \right] \right] \quad (6)$$

2.2 Delimiting the cell into zones

As mentioned earlier, PF falls short from realizing exact fairness in a realistic environment where users experience heterogeneous fading. To be more precise, in order for all users in the PF algorithm to get access to the channel the same asymptotical fraction of time and get the same average power independently of their distance to the BS, two main assumptions must be met:

- *Assumption I*: the fading must be homogeneous and independent among users.
- *Assumption II*: either the instantaneous rate must scale linearly with the instantaneous SNR or all users must have the same average SNR.

Unfortunately, in practice, these assumptions are not valid: *first*, users do not experience the same type of fading which is a very complex phenomenon that varies widely across users. As a result, while the random variables representing the fading effects are independent among users, they are not identically distributed. The principal impact of this lack of homogeneity is an unfair distribution of slots amid active users: users with the most variable distributions (typically those who are the furthest away from the BS) receive the least amount of slots. *Second*, the linearity between the feasible rate and SNR is too optimistic except for users with low SNR (again, typically for users far from the BS) and users are unlikely to have the same mean SNR.

Our target is to define the previously introduced zones in such a way that the two cited assumptions become more realistic inside each zone. For that reason, we will follow the approach taken in [2] as it serves well our purposes (although the reasons why the zones were introduced in the cited paper are completely different). We take a path loss exponent $\beta=4$ as we consider urban environments. The random variables x_k are exponentially distributed (with unit mean) as we consider Rayleigh fading.

We define the zones as follows:

2.2.1 Zone 0

Users in this zone are those who get the maximum peak rate C_0 with probability equal to 0.95. Hence, from (5), r_1 is given by:

$$\begin{aligned} \mathbb{P}(R_{k,0}(r_1, t) = C_0) &= 0.95 \Rightarrow \\ \mathbb{P}\left(\left(\frac{r^*}{r_1}\right)^4 \cdot x_k \cdot K_0 > 1\right) &= 0.95 \Rightarrow \\ e^{-\left(\frac{r^*}{r_1}\right)^4} &= 0.95 \Rightarrow r_1 = (-\ln(0.95))^{1/4} r^* \end{aligned}$$

The mean rate of a user k in Zone 0 is then $C_{k,0}(r) \approx C_0$.

2.2.2 Zone 2

Users in this zone are those who do not get C_0 with probability equal to 0.95. Hence, from (5), r_2 is given by:

$$\begin{aligned} \mathbb{P}(R_{k,2}(r_2, t) \neq C_0) &= 0.95 \Rightarrow \\ \mathbb{P}\left(\left(\frac{r^*}{r_2}\right)^4 \cdot x_k \cdot K_2 < 1\right) &= 0.95 \Rightarrow \\ e^{-\left(\frac{r^*}{r_2}\right)^4 \cdot \frac{1}{K_2}} &= 0.05 \Rightarrow r_2 = (-\ln(0.05)K_2)^{1/4} r^* \end{aligned}$$

To obtain the value of r_2 , we need to determine the value of K_2 . For that, we consider hexagonal networks where the interference suffered by a user in a given cell is almost utterly generated by the 6 neighbouring BSs. This assumption is fairly valid in an urban environment where the path loss is at least equal to 4. An approximation of this interference is given by [2]:

$$I(r) = P \cdot [\gamma(2\Re - r) + 2 \cdot \gamma \left(\sqrt{(\Re - r)^2 + 3\Re^2} \right) + \gamma(2\Re + r) + 2 \cdot \gamma \left(\sqrt{(\Re + r)^2 + 3\Re^2} \right)]$$

We take $\Re = 2 \cdot r^*$ as larger values of \Re induce very small rates at the border of the cell [2].

If $\eta \ll I \Rightarrow \frac{\eta+I_0}{\eta+I_2} \rightarrow \frac{I_0}{I_2} \approx \frac{I(r_1)}{I(\Re)} \approx 0.31$ and if $\eta \gg I \Rightarrow \frac{\eta+I_0}{\eta+I_2} \rightarrow 1$. Due to the fact that $\frac{\eta+I_0}{\eta+I_2}$ is an increasing function in η , we have that $0.31 \leq \frac{\eta+I_0}{\eta+I_2} \leq 1$. Knowing that $\frac{\sigma_0}{\sigma_2} \approx 2.51$ [1], we deduce that $0.78 \leq K_2 \leq 2.51$.

The mean rate of a user k in Zone 2 is then $C_{k,2}(r) \approx C_0 \left(\frac{r^*}{r}\right)^4 K_2$ and the distribution of the feasible rates is approximately that of $C_{k,2}(r) \cdot x_k(t)$.

2.2.3 Zone 1

Users in this zone are consequently those who get their maximum peak rate C_0 with non-negligible probability. We assume that, for this intermediate zone, the distribution of the feasible rate is approximately the same for all users:

$$x'_k(t) \cdot C_{k,1} = \int_{r_1}^{r_2} C_0 \cdot \min\left[\left(\frac{r^*}{r}\right)^4 \cdot x_k(t) \cdot K_1, 1\right] \cdot \frac{2rdr}{r_2^2 - r_1^2} \quad (7)$$

with x'_k being the unit mean random variable representing the variations, due to fading, around the mean rate $C_{k,1}$ of user k in Zone 1. This approximation is made in order to obtain homogeneous fading in this zone.

From (7), we get the following:

$$C_1 = \int_{r_1}^{r_2} \mathbb{E}[\min\left[\left(\frac{r^*}{r}\right)^4 K_1 \cdot x_k(t), 1\right]] \cdot \frac{2rdr}{r_2^2 - r_1^2}$$

with $\mathbb{E}[\min[(\frac{r^*}{r})^4 K_1 \cdot x_k(t), 1]] = \frac{1 - e^{-(\frac{r^*}{r})^4 \frac{1}{K_1}}}{(\frac{r^*}{r})^4 \frac{1}{K_1}}$.

To compute K_1 , we proceed as for Zone 2. We take $K_2 = 1.7$ and thus $r_2 \approx 1.5 \cdot r^*$. If $\eta \ll I \Rightarrow \frac{\eta+I_0}{\eta+I_1} \rightarrow \frac{I_0}{I_1} \approx \frac{I(r_1)}{I(r_2)} \approx 0.71$ and if $\eta \gg I \Rightarrow \frac{\eta+I_0}{\eta+I_1} \rightarrow 1$. We deduce that $0.71 \leq \frac{\eta+I_0}{\eta+I_1} \leq 1$. Having that $\frac{\sigma_0}{\sigma_1} \approx 1.3$ [1], we finally have that $0.92 \leq K_1 \leq 1.3$.

We conclude that despite the heterogeneity of fading all over the cell, by dividing it into the three previous zones, we can fairly assume the homogeneity of fading within each zone, required to obtain a "fair" scheduling (applied among users of the same zone). Hence, *Assumption I* holds in our proposed model. Besides, *Assumption II* is valid in our model for Zones 0 and 2, and better approximated for Zone 1.

In summary, for user k in Zone z , we can write the feasible rate as $R_{k,z} = C_{k,z} \cdot X_{k,z}$ by defining $X_{k,z}$ as being the variations due to fading (of unit mean) around the mean rate $C_{k,z}$. More explicitly, we have for Zone 0 $X_{k,0} = 1$, for Zone 1 $X_{k,1} = x'_k$ (which is a function of x_k and is given in details in Appendix 6.1) and for Zone 2 $X_{k,2} = x_k$.

Remark: There exist different methods which enable the BS to localize users and hence to classify them into different zones [8].

3 Analytical study of PF Vs. WPF

The PF algorithm is thoroughly studied in [5] in the case of homogeneous fading. We start by reminding some results from the cited paper that we use in our analysis.

At time slot t , PF schedules the user with the highest feasible rate relative to its current average throughput, i.e.,

$$user\ k^* = \operatorname{argmax}_k [\frac{R_k(t)}{T_k(t)}]$$

with $T_k(t)$ being the exponentially smoothed throughput given by:

$$T_k(t+1) = (1 - \frac{1}{\tau}) \cdot T_k(t) + \frac{1}{\tau} \cdot R_k(t) \cdot \mathbb{1}_{user(t)=k} \quad (8)$$

where $\mathbb{1}_{user(t)=k}$ is the indicator function which equals 1 if user k was chosen at time slot t and 0 otherwise. τ is a time constant that captures the time-scales of the PF scheduler.

Since the random variables representing the fading are i.i.d, we have that:

$$T_k(t) = C_k \cdot U_k(t)$$

where C_k is the mean rate of user k and U_k are identically distributed random variables.

If $\frac{1}{\tau} \rightarrow 0$, then:

$$T_k \rightarrow C_k \cdot \frac{g(n)}{n} \quad (9)$$

where n is the total number of active users and $g(n) = \mathbb{E}[\max(X_1, \dots, X_n)]$ is the PF scheduling gain. In practice, τ has large values as this offers the opportunity of waiting a long time before scheduling a user when its channel quality is maximal: the scheduler is then expected to better exploit multi-user diversity. Hence, we adopt formula (9) in our analysis.

Next, in subsection 3.1, we analyse the WPF scheduler. Since there is not an exhaustive study of PF with heterogeneous fading in the literature, we propose, in subsection 3.2, an approximate analysis for PF in a realistic environment. In subsection 3.3, we corroborate the validity of our results through simulation.

3.1 The WPF Scheduler

In our model, an independent PF scheduler is applied among users belonging to the same zone and thus experiencing homogeneous fading. Therefore, we can adopt the result in (9) for each zone. Namely, in Zone z , the exponentially smoothed throughput of user k is given by:

$$T_{k,z} \rightarrow C_{k,z} \cdot \frac{g_z(n_z)}{n_z} \quad (10)$$

where n_z is the total number of active users in Zone z and $g_z(n_z) = \mathbb{E}[\max(X_{1,z}, \dots, X_{n_z,z})]$.

3.1.1 Average Rate

The average rate of a user k belonging to Zone z is then:

$$\chi_{k,z,WPF} = \frac{C_{k,z}}{n_z} \cdot g_z(n_z) \cdot \mathbb{P}(\alpha_z) \quad (11)$$

with event $\alpha_z = \{\text{Zone } z \text{ is served}\}$.

Proof: The average rate of a user k in Zone z served according to WPF is:

$$\begin{aligned} & \mathbb{E}[R_{k,z} \cdot \mathbb{1}\{\frac{R_{k,z}}{T_{k,z}} = \max_{l=1..n_z} \frac{R_{l,z}}{T_{l,z}}\} \cdot \mathbb{1}\{\alpha_z\}] \\ &= C_{k,z} \cdot \mathbb{E}[X_{k,z} \cdot \mathbb{1}\{\frac{X_{k,z}}{\frac{g_z(n_z)}{n_z}} = \max_{l=1..n_z} \frac{X_{l,z}}{\frac{g_z(n_z)}{n_z}}\} \cdot \mathbb{P}(\alpha_z)] \\ &= C_{k,z} \cdot \mathbb{E}[X_{k,z} \cdot \mathbb{1}\{X_{k,z} = \max_{l=1..n_z} X_{l,z}\} \cdot \mathbb{P}(\alpha_z)] \end{aligned}$$

Since the random variables $X_{k,z}$ are i.i.d, we obtain the following:

$$\frac{C_{k,z}}{n_z} \cdot \mathbb{E}[\max(X_{1,z}, \dots, X_{n_z,z})] \cdot \mathbb{P}(\alpha_z)$$

We denote by $G_z(n_z)$ the scheduling gain in Zone z , defined as the ratio of what the user receives as compared to a simple RR (Round Robin) scheduling in this zone and given by:

$$G_z(n_z) = \frac{\chi_{k,z,WPF}}{\frac{C_{k,z}}{n_z}} = g_z(n_z) \cdot \mathbb{P}(\alpha_z)$$

The average rate per user in Zone z is then:

$$\chi_{z,WPF} = \frac{C_z \cdot G_z(n_z)}{n_z}$$

where C_z is the mean rate perceived by a user in Zone z . Namely, C_0 in Zone 0, C_1 in Zone 1 and $C_2 = C_0 K_2 \cdot \int_{r_2}^R (\frac{r^*}{r})^4 \frac{2rdr}{R^2 - r^2}$ in Zone 2.

3.2 The PF Scheduler

We analyse a model where PF selects a user among all users present in the cell while adopting for the exponentially smoothed throughput the value taken by formula (10). We make this approximation in order to have a tractable model by supposing that formula (10) remains valid for a user in a given Zone z under plain PF. We prove the validity of this assumption through simulation in subsection 3.3.

3.2.1 Average Rate

The average rate of a user k belonging to Zone z is then:

$$\chi_{k,z,PF} = \frac{C_{k,z}}{n_z} \cdot \mathbb{E}[Z_z \cdot \mathbb{1}\{\frac{Z_z \cdot n_z}{g_z(n_z)} > \frac{Z_j \cdot n_j}{g_j(n_j)}, \forall j \neq z\}] \quad (12)$$

with $Z_j = \max\{X_{1,j}, \dots, X_{n_j,j}\}$.

Proof: The average rate of a user k in Zone z served according to PF is:

$$\begin{aligned} & \mathbb{E}[R_{k,z} \cdot \mathbb{1}\{\frac{R_{k,z}}{T_{k,z}} = \max_{j=0,1,2} \max_{l=1..n_j} \frac{R_{l,j}}{T_{l,j}}\}] \\ &= C_{k,z} \cdot \mathbb{E}[X_{k,z} \cdot \mathbb{1}\{\frac{X_{k,z}}{\frac{g_z(n_z)}{n_z}} = \max_j \max_l \frac{X_{l,j}}{\frac{g_j(n_j)}{n_j}}\}] \\ &= C_{k,z} \cdot \mathbb{E}[X_{k,z} \cdot \mathbb{1}\{\frac{X_{k,z}}{\frac{g_z(n_z)}{n_z}} > \frac{Z_j}{\frac{g_j(n_j)}{n_j}}, \forall j \neq z\}] \\ & \quad \mathbb{1}\{\frac{X_{k,z}}{\frac{g_z(n_z)}{n_z}} = \frac{Z_z}{\frac{g_z(n_z)}{n_z}}\}] \\ &= C_{k,z} \cdot \mathbb{P}(A) \cdot \mathbb{E}[X_{k,z} \cdot \mathbb{1}\{\frac{X_{k,z}}{\frac{g_z(n_z)}{n_z}} > \frac{Z_j}{\frac{g_j(n_j)}{n_j}}, \forall j \neq z\} | A] \end{aligned}$$

with event $A = \{X_{k,z} = Z_z\}$. The random variables $X_{k,z}$ being i.i.d, formula (12) follows.

We deduce from (12) the scheduling gain in Zone z denoted by $G_z(n_0, n_1, n_2)$ and defined as the ratio of what the user receives as compared to a simple RR scheduling (the detailed formulae are found in Appendix 6.1):

$$G_z(n_0, n_1, n_2) = \frac{\chi_{k,z,PF}}{\frac{C_{k,z}}{n_z}}$$

The average rate per user in Zone z is then:

$$\chi_{z,PF} = \frac{C_z \cdot G_z(n_0, n_1, n_2)}{n_z}$$

3.2.2 Access Probability

To evaluate the impact of heterogeneous fading on the access probability when applying PF to the whole cell, we define the following probabilities (the detailed formulae are found in Appendix 6.2):

$$a = \mathbb{P}(Z_2 \cdot \frac{C_2}{T_2} > Z_1 \cdot \frac{C_1}{T_1}), \quad b = \mathbb{P}(Z_2 \cdot \frac{C_2}{T_2} > \frac{C_0}{T_0}) \quad \text{and} \\ c = \mathbb{P}(Z_1 \cdot \frac{C_1}{T_1} > \frac{C_0}{T_0}).$$

Therefore, the probability to serve a user in Zone 2 is $P_2 = \frac{a \cdot b}{n_2}$, the probability to serve a user in Zone 1 is $P_1 = \frac{(1-a) \cdot c}{n_1}$ and the probability to serve a user in Zone 0 is $P_0 = \frac{(1-b) \cdot (1-c)}{n_0}$.

If we had homogeneous fading all over the cell, the probability of a user k to be selected would be the same for all users in the cell and is given by $P = \mathbb{P}(X_k = \max(X_1, \dots, X_n)) = \frac{1}{n}$.

3.3 Numerical results

We present in this section our numerical experiments performed to illustrate the previous results. The number of users in each zone is fixed and equal to 10 ($n_0 = n_1 = n_2 = 10$ and thus $n = 30$). We take $C_0 = 1$, $r^* = 1$, $K_2 = 1.7$ and $K_1 = 1$. As a result, we get $r_1 \approx 0.5$ and $r_2 \approx 1.5$. Rayleigh Fading is considered corresponding to an exponential distribution of the process $x_k(t)$.

Users are served according to PF and according to our WPF that comprises three different scenarios:

- **WPF(1,1,1)** where slots are distributed fairly across zones ($\mathbb{P}(\alpha_0) = \mathbb{P}(\alpha_1) = \mathbb{P}(\alpha_2) = 1/3$).
- **WPF(1,1,4)** where, in order to favor far users who are penalized by their small rates, Zone 2 is given four

more slots than Zone 1 and Zone 0 ($\mathbb{P}(\alpha_0) = \mathbb{P}(\alpha_1) = 1/6$ and $\mathbb{P}(\alpha_2) = 2/3$).

- **WPF(4,2,1)** where, in order to increase the total capacity, Zone 0 is given twice as many slots as Zone 1 and the latter is given twice as many slots as Zone 2 ($\mathbb{P}(\alpha_0) = 4/7$, $\mathbb{P}(\alpha_1) = 2/7$ and $\mathbb{P}(\alpha_2) = 1/7$).

In all experiments, we determine the average rate per user and display the results for users belonging to the same zone for PF and for WPF. We compare the results obtained by simulation to those obtained numerically from $\chi_{z,PF}$ and $\chi_{z,WPF}$.

Results are shown in Table 1 and indicate that the analytical formulae provide highly accurate estimates of simulation results and thus our proposed approximation for PF is valid. As for performance, we see, in WPF(1,1,1), that there is a conservation in the mean rate in Zone 0 and Zone 1 in comparison with PF, contrary to Zone 2 where a slight degradation is witnessed. This loss is compensated for by the increase in the probability to access the channel for users in Zone 2. Indeed, from subsection 3.2.2, P_0 and P_1 are roughly equal to $1/30$ (as in the homogeneous case from P) while P_2 is approximately equal to 0.017 (whereas P gives $1/30$ for the homogeneous case). This means that with heterogeneous fading, the PF algorithm favors close users who are scheduled about twice as often as far users. Yet, in WPF, if the system wants to share resources fairly between users, it can give one time slot to each zone successively as it is done in WPF(1,1,1) where $\mathbb{P}(\alpha_z) = 1/3$. In other words, the probability to choose a user will be $1/3 \cdot 1/10$ (recall that users in the same Zone z have the same probability $\frac{1}{n_z}$ to be selected) exactly as in the ideal homogeneous system. Far users will then have the same chance to be scheduled as close users and hence the lost equity in the distribution of slots among users will be restored.

Furthermore, the system can now be entirely controlled as a significant flexibility in the allocation of resources is introduced:

- If the system seeks to realize a fair rate sharing, it can give more slots to far users who are disadvantaged as compared to nearer users due to their small feasible rates, as for instance in WPF(1,1,4) where users in Zone 2 perceive a gain of 50% in comparison with PF.
- However, if the system wants to maximize the overall throughput without causing the starvation of distant users (as it is done in the Max SNR scheduler that always serves the user with highest rate [4]), it can give more slots to close users than to far away users. Indeed, in WPF(4,2,1), users in Zone 0 perceive a gain

Zone 0	PF	WPF(1,1,1)	WPF(4,2,1)	WPF(1,1,4)
NUM	0.034	0.034	0.065	0.0167
SIM	0.033	0.033	0.061	0.0167
Zone 1	PF	WPF(1,1,1)	WPF(4,2,1)	WPF(1,1,4)
NUM	0.033	0.0297	0.026	0.0165
SIM	0.033	0.0315	0.0274	0.0160
Zone 2	PF	WPF(1,1,1)	WPF(4,2,1)	WPF(1,1,4)
NUM	0.013	0.010	0.0042	0.0195
SIM	0.012	0.011	0.0045	0.0208

Table 1. Average Rate per user

of 76% as compared to PF leading to a gain in global rate of approximately 20%.

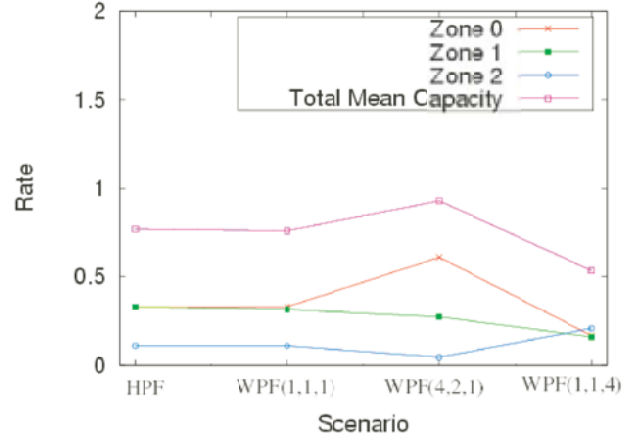


Figure 1. Mean Rate per Zone and Total Mean Capacity

We graphed in Figure 1 the mean rate per Zone and the Total Mean capacity of the cell for PF and for WPF (from simulations results of Table 1). The variations of curves with the variation of scenarios highlight the mentioned trade-off between fairness and efficiency and clearly show how it is now fully controlled.

4 Dynamic Model of WPF

We deduce from formula (11) in subsection 3.1 that each zone, in our WPF system, behaves like a Generalized Processor Sharing (GPS) queue whose service rate is $\frac{G_z(n_z)}{n_z}$ (we refer here to the definition of GPS given in [3] which denotes a PS queue where the rate at which users are served is an arbitrary positive function of the number of users). In particular, Zone 0 behaves like a simple PS

system owing to the fact that $g_0(n_0) = 1$. Whereas with heterogeneous fading, PF loses the symmetry properties of the GPS discipline because the rate at which users are served in a given Zone z is a function of the number of users in all zones as we can see from formula (12). Therefore, our proposed model, contrary to PF, is fully tractable in the dynamic case.

We assume that data flows arrive as a Poisson process of intensity $\lambda \cdot ds$ in any area of surface ds . Flow sizes are i.i.d and S is the corresponding random variable. We denote by $\rho = \lambda \cdot \mathbb{E}[S]$ the traffic load and by $d\rho(r) = \rho \cdot 2\pi r dr$ the traffic load generated by users whose distance to the BS ranges between r and $r + dr$. Such a system has the well-known insensitivity property which means that performance depends mainly on the load factor (and on the maximum number of users in presence of an admission control policy) and not on the distribution of the flow size which is continually changing given the ever varying nature of data applications. Thus, our scheduling approach, in addition to making the system more flexible, simplifies dimensioning.

4.1 Analytical Model

From [3], we obtain the stationary distribution of the number of users in Zone z :

$$\pi_z(x) = \frac{\prod_{i=1}^x \frac{\rho_z}{G_z(i)}}{\sum_{k=0}^{n_z} \prod_{i=1}^k \frac{\rho_z}{G_z(i)}}$$

where ρ_z is the load in Zone z and n_z is the maximum number of admitted users in this zone.

More explicitly, we have the following:

For Zone 0, $\rho_0 = \int_{r_0}^{r_1} \frac{d\rho(r)}{C_0} = \frac{\rho\pi r_1^2}{C_0}$.

For Zone 1, $\rho_1 = \int_{r_1}^{r_2} \frac{d\rho(r)}{C_1} = \frac{\rho\pi(r_2^2 - r_1^2)}{C_1}$.

For Zone 2, $\rho_2 = \int_{r_2}^{\mathcal{R}} \frac{d\rho(r)}{C_0 \cdot (\frac{r}{r_2})^\beta} = \frac{\rho\pi(\mathcal{R}^{\beta+2} - r_2^{\beta+2})}{C_0 \cdot r_2^{\beta+2} \cdot (\beta+2)}$.

Using Little's law, we find that the flow throughput $Th_{i,z}$ of user i in Zone z , defined as the ratio of the mean flow size $\mathbb{E}[S]$ to the mean flow duration, is given by:

$$Th_{i,z} = C_{i,z} \cdot \frac{\rho_z \cdot (1 - B_z)}{\mathbb{E}[n_z]}$$

where $B_z = \pi_z(x = n_z)$ is the blocking probability and $\mathbb{E}[n_z] = \sum_{i=1}^{n_z} i \cdot \pi_z(i)$ is the mean number of active users in Zone z .

4.2 Simulation Results

We present here numerical experiments that we performed to illustrate the above results. We consider a system

where users initiate file transfer requests as a Poisson process of intensity $\lambda\pi\mathcal{R}^2$ and traffic demand is uniformly distributed in the cell. Rayleigh Fading is considered corresponding to an exponential distribution of the process $x_k(t)$. At most 10 users are admitted simultaneously in each zone to guarantee a minimum rate of $\frac{C(\mathcal{R})}{30}$. Guaranteeing a minimum rate is a QoS notion appropriate for non-real time users. New transfers generated in a zone where there are already 10 transfers in progress are blocked and lost. Flow sizes are independent and exponentially distributed with mean equal to 2500 Kbits (which means equal to 1 for $C_0 = 1$).

We determine the throughput per user and display the average throughput for users belonging to the same zone. Users are served according to our WPF(1,1,1). The simulation results obtained are compared to the normalized throughput $\frac{Th_{i,z}}{C_{i,z}}$ of subsection 4.1.

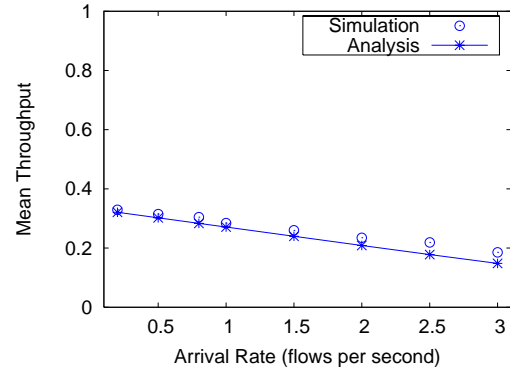


Figure 2. Average Throughput for Zone 0

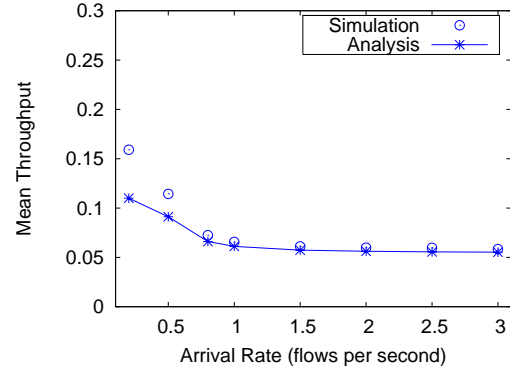


Figure 3. Average Throughput for Zone 1

Figures 2, 3 and 4 depict the mean throughput as a function of the file arrival rate respectively for Zone 0, Zone 1

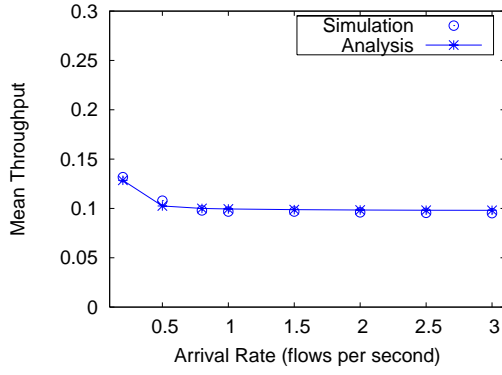


Figure 4. Average Throughput for Zone 2

and Zone 2. We can see how the analysis and simulation curves are nearly indistinguishable for Zone 0 and Zone 2 because both of these zones behave perfectly like a GPS system. Yet, for Zone 1, the analytical model gives less accurate predictions because the approximation made for this zone is not totally valid since it still experiences heterogeneous fading in a realistic environment: the closest users in Zone 1 have almost always fixed rates while the farthest users in the same zone have continuously varying rates. Nonetheless, we can fairly consider that the proposed analytical model approximates very well the behaviour of our WPF scheduler.

5 Conclusion

In this paper, we propose a new scheduling approach for new generation cellular networks. The approach is based on the partitioning of the cell into zones in such a way that inside each zone the fading and SNR are quite homogenous. We propose a hierarchical scheduler, named WPF, which, at a first level, serves the zones in a WRR fashion and, at a second level, serves users within each zone with independent PF schedulers. Our analytical study (validated by simulation) shows how our approach allows controlling the trade off between global efficiency and fairness, a capability that PF lacks for. Moreover, we propose an analytical model for the PF algorithm under realistic heterogeneous fading conditions that we validate by simulation. An analysis of WFP under dynamic traffic conditions is also provided.

6 Appendix

We denote by $u = \frac{1}{K_1} \cdot (\frac{r_1}{r_*})^4$, $v = \frac{1}{K_1} \cdot (\frac{r_2}{r_*})^4$, $K = \frac{C_0 K_1}{C_1}$ and $Z_j = \max\{X_{1,j}, \dots, X_{n_j,j}\}$.

6.1 Computing $G_z(n_0, n_1, n_2)$:

From (7), we know that $X_{k,1} = x'_k$ is a function of $X_{k,2} = x_k$. Hence, we obtain the following:

$$x'_k = f(x_k) = \begin{cases} \frac{K}{\sqrt{uv}} \cdot x_k & \text{if } x_k \leq u \\ \frac{K(2\sqrt{x_k} - \sqrt{u} - \frac{x_k}{\sqrt{v}})}{(\sqrt{v} - \sqrt{u})} & \text{if } u \leq x_k \leq v \\ K & \text{if } x_k \geq v \end{cases}$$

The random variables $X_{k,2}$ being exponentially distributed, we have that:

$$\begin{aligned} g_2(n_2) &= \mathbb{E}[\max(X_{1,2}, \dots, X_{n_2,2})] = \int_0^\infty \mathbb{P}(Z_2 > s) ds \\ &= \int_0^\infty 1 - (1 - e^{-s})^{n_2} ds = \sum_{i=1}^{n_2} \frac{1}{i} \end{aligned}$$

$$\begin{aligned} g_1(n_1) &= \mathbb{E}[\max(X_{1,1}, \dots, X_{n_1,1})] = \int_0^\infty \mathbb{P}(Z_1 > s) ds \\ &= \int_0^{f(v)} 1 - (1 - e^{-f^{-1}(s)})^{n_1} ds \end{aligned}$$

with

$$f^{-1}(s) = \begin{cases} \frac{\sqrt{uv}}{K} \cdot s & \text{if } 0 \leq s \leq f(u) \\ (\sqrt{v} - [v - (\sqrt{uv} + s \frac{\sqrt{v}(\sqrt{v} - \sqrt{u})}{K})]^{1/2})^2 & \text{if } f(u) \leq s \leq f(v) \end{cases}$$

And $g_0(n_0) = 1$ (because $X_{k,0} = 1$).

We denote by $h_2(z)$ and $h_1(z)$ the density functions of respectively Z_2 and Z_1 which are given by ($Z_0 = 1$):

$$\begin{aligned} h_2(z) &= n_2 e^{-z} (1 - e^{-z})^{n_2-1} \\ h_1(z) &= \begin{cases} \frac{n_1 e^{-f^{-1}(z)} (1 - e^{-f^{-1}(z)})^{n_1-1}}{f' \circ f^{-1}(z)} & \text{if } 0 \leq z < f(v) \\ \frac{(1 - (1 - e^{-v})^{n_1}) \cdot \delta(z - f(v))}{(1 - (1 - e^{-v})^{n_1}) \cdot \delta(z - f(v))} & \text{if } z \geq f(v) \end{cases} \end{aligned}$$

The gain in Zone 2 is $G_2(n_0, n_1, n_2) =$

$$\begin{aligned} &\mathbb{E}[Z_2 \cdot \mathbf{1}\{\frac{Z_2 n_2}{g_2(n_2)} \geq \frac{Z_1 n_1}{g_1(n_1)}\} \cdot \mathbf{1}\{\frac{Z_2 n_2}{g_2(n_2)} \geq n_0\}] = \\ &\mathbb{E}[Z_2 \cdot \mathbf{1}\{\frac{Z_2 n_2}{g_2(n_2)} \geq \frac{Z_1 n_1}{g_1(n_1)} \geq n_0\}] + \\ &\mathbb{E}[Z_2 \cdot \mathbf{1}\{\frac{Z_2 n_2}{g_2(n_2)} \geq n_0 \geq \frac{Z_1 n_1}{g_1(n_1)}\}] = \\ &\int_{\frac{g_1(n_1)n_0}{n_1}}^{\frac{g_2(n_2)n_1}{g_1(n_1)n_2}} h_1(z_1) \int_{\frac{g_2(n_2)n_1}{g_1(n_1)n_2}}^{\infty} z_2 h_2(z_2) dz_2 dz_1 + \\ &\mathbb{P}(Z_1 \leq \frac{g_1(n_1)n_0}{n_1}) \int_{\frac{g_2(n_2)n_0}{n_2}}^{\infty} z_2 h_2(z_2) dz_2 \end{aligned}$$

The gain in Zone 1 is $G_1(n_0, n_1, n_2) =$

$$\begin{aligned} & \mathbb{E}[Z_1 \cdot \mathbf{1}\{\frac{Z_1 n_1}{g_1(n_1)} \geq \frac{Z_2 n_2}{g_2(n_2)}\} \cdot \mathbf{1}\{\frac{Z_1 n_1}{g_1(n_1)} \geq n_0\}] = \\ & \mathbb{E}[Z_1 \cdot \mathbf{1}\{\frac{Z_1 n_1}{g_1(n_1)} \geq \frac{Z_2 n_2}{g_2(n_2)} \geq n_0\}] + \\ & \mathbb{E}[Z_1 \cdot \mathbf{1}\{\frac{Z_1 n_1}{g_1(n_1)} \geq n_0 \geq \frac{Z_2 n_2}{g_2(n_2)}\}] = \\ & \int_{\frac{g_2(n_2)n_0}{n_2}}^{\infty} h_2(z_2) \int_{\frac{g_1(n_1)n_2}{g_2(n_2)n_1} \cdot z_2}^{\infty} z_1 h_1(z_1) dz_1 dz_2 + \\ & \mathbb{P}(Z_2 \leq \frac{g_2(n_2)n_0}{n_2}) \int_{\frac{g_1(n_1)n_0}{n_1}}^{\infty} z_1 h_1(z_1) dz_1 \end{aligned}$$

The gain in Zone 0 is $G_0(n_0, n_1, n_2) =$

$$\begin{aligned} & \mathbb{E}[\mathbf{1}\{\frac{Z_1 n_1}{g_1(n_1)} \leq n_0\} \cdot \mathbf{1}\{\frac{Z_2 n_2}{g_2(n_2)} \leq n_0\}] = \\ & \mathbb{P}(Z_2 \leq \frac{g_2(n_2)n_0}{n_2}) \cdot \mathbb{P}(Z_1 \leq \frac{g_1(n_1)n_0}{n_1}) \end{aligned}$$

6.2 Computing the probabilities a , b & c :

$$\begin{aligned} a &= \mathbb{P}(\frac{Z_2 C_2}{T_2} > \frac{Z_1 C_1}{T_1}) = \mathbb{P}(Z_1 < Z_2 \frac{g_1(n_1)n_2}{g_2(n_2)n_1}) \\ &= \int_0^{\infty} \mathbb{P}(z_1 < s \cdot \frac{g_1(n_1)n_2}{g_2(n_2)n_1}) \cdot h_2(s) ds \\ &= \int_0^{\frac{g_2(n_2)n_1}{g_1(n_1)n_2} \cdot f(v)} (1 - e^{-f^{-1}(s \frac{g_1(n_1)n_2}{g_2(n_2)n_1})})^{n_1} h_2(s) ds + \\ & \int_{\frac{g_2(n_2)n_1}{g_1(n_1)n_2} \cdot f(v)}^{\infty} h_2(s) ds \\ b &= \mathbb{P}(\frac{Z_2 C_2}{T_2} > \frac{C_0}{T_0}) \\ &= \mathbb{P}(Z_2 > \frac{g_2(n_2)n_0}{n_2}) = \int_{\frac{g_2(n_2)n_0}{n_2}}^{\infty} h_2(s) ds \\ c &= \mathbb{P}(\frac{Z_1 C_1}{T_1} > \frac{C_0}{T_0}) \\ &= \mathbb{P}(Z_1 > \frac{g_1(n_1)n_0}{n_1}) = \int_{\frac{g_1(n_1)n_0}{n_1}}^{\infty} h_1(s) ds \end{aligned}$$

References

- [1] P. Bender. Cdma/hdr: A bandwidth-efficient high-speed wireless data service for nomadic users. *IEEE Communications Magazine*, 38(7):70–77, July 2000.
- [2] T. Bonald. Wireless downlink data channels: User performance and cell dimensioning. *MobiCom'03*, 2003.

- [3] J. W. Cohen. The multiple phase service network with generalized processor sharing. *Acta Informatica*, 12:245–284, 1979.
- [4] J. Holtzman. Asymptotic analysis of proportional fair algorithm. *Proceedings of IEEE*, 2001.
- [5] H. Kushner. Convergence of proportional-fair sharing algorithms under general conditions. *IEEE Communications Magazine*, 3(4):1250–1259, July 2004.
- [6] R. Love. High-speed downlink packet access performance. *Proceedings of IEEE VTC Spring*, 2001.
- [7] A. Viterbi. *CDMA: Principles of Spread Spectrum Communication*. Addison-Wesley, 1995.
- [8] Y. Zhao. Standardization of mobile phone positioning for 3g systems. *IEEE Communications Magazine*, 40(7):108–116, July 2002.