

Asymptotic Properties of Proportional-Fair Sharing Algorithms

Harold J. Kushner*
Division of Applied Mathematics
Brown University, Providence RI 02912

Philip A. Whiting
Room 2C-317, Bell Labs, Lucent Technologies, 600 Mountain Ave,
Murray Hill NJ 07974-0636
Tel: +1 908 508 1593 Fax:+1 908 508 1593
email: pwhiting@research.bell-labs.com

May 30, 2002

Note: The current draft is only a brief outline, with no proofs: More detail will be in the final paper.

Abstract

We are concerned with the allocation of channel or transmitter resources for time varying mobile communications. There are many users who are competing to transmit data over the resource. Time is divided into small scheduling intervals, and information on the channel rates for the various users is available at the start of the intervals. Since the rates vary randomly, there is a conflict at any time between fully exploiting the channel (by selecting the user with the highest current rate) and being fair (giving attention to users with poor rates, to assure a fair throughput for them). The Proportional Fair Scheduler (PFS) of the Qualcomm High Data Rate (HDR) system and related algorithms are designed to deal with such conflicts. There is little analysis available for such systems and our aim is to put them on a sure mathematical footing and analyze their behavior. Such algorithms are of the stochastic approximation type and results of stochastic approximation are used to analyze the long term properties of this class. The limiting behavior of the throughputs converges to the solution of an ordinary differential equation (a mean ODE), which is akin to a mean flow. The ODE has a unique equilibrium and it is optimal in the sense that it optimizes a concave utility function. The results depend on the fact that the mean ODE has a special form that arises

*Partially supported by Army Research Office Contract DAAD19-00-1-0549 and National Science Foundation Grants ECS 9979250 and ECS 0097447.

Report Documentation Page

Form Approved
OMB No. 0704-0188

Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.

1. REPORT DATE 30 MAY 2002		2. REPORT TYPE		3. DATES COVERED 00-05-2002 to 00-05-2002	
4. TITLE AND SUBTITLE Asymptotic Properties of Proportional-Fair Sharing Algorithms				5a. CONTRACT NUMBER	
				5b. GRANT NUMBER	
				5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S)				5d. PROJECT NUMBER	
				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Brown University, Division of Applied Mathematics, 182 George Street, Providence, RI, 02912				8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)				10. SPONSOR/MONITOR'S ACRONYM(S)	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release; distribution unlimited					
13. SUPPLEMENTARY NOTES					
14. ABSTRACT					
15. SUBJECT TERMS					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES 9	19a. NAME OF RESPONSIBLE PERSON
a. REPORT unclassified	b. ABSTRACT unclassified	c. THIS PAGE unclassified			

in problems with certain types of repeated stochastic games with competitive behavior. There are a large family of such algorithms, each member corresponding to a concave utility function. Thus, PFS is not simply ad-hoc, but actually corresponds to a reasonable maximization problem. There are extensions to multiple antenna and frequency systems. Also, the infinite backlog assumption can be dropped and the data is allowed to arrive at random.

1 Introduction

HDR provides data connection for a set of users via a common shared downlink, onto which user transmissions are scheduled. Access to the link is given one user at a time for a fixed duration time slot of about 1.67 ms. The rates of transmission (based, say, on the response to pilot signals) for the various users are known at the beginning of the slot. Since the time between measurement and prediction is short, fairly accurate rate predictions (taken in this paper to be exact) can be made. Scheduling decisions can take into account Rayleigh fading with a frequency of a few tens of Hertz. A problem for HDR is how to “fair share” the slots - if the highest declared rate is always chosen, users with high SNR will consistently be selected, starving low SNR users.

The *Proportional Fair scheduler*, selects mobiles by comparing their current rates with their past averaged throughputs and selecting the mobile with highest relative throughput as in (6). The algorithm proposed by Qualcomm performs this sharing by comparing the declared rates with the users long run throughputs at the n th timeslot and scheduling the user with the highest relative throughput as in (6). PFS originates in the allocation of connections over multiple links on the Internet [6]. In wireless, it allows “scheduling according to the peak channel fluctuations.” There has been little rigorous analysis of such algorithms, and our aim is to put them on a solid mathematical foundation.

The algorithms fall into the area that is known as stochastic approximation in which there is a wide range of useful results [8]. The main one used here is convergence of the sample paths of the throughputs to trajectories of a corresponding dynamical system, the characteristics of which can be obtained from the rate distribution; see Theorem 3.1.

It is difficult to solve these equations explicitly, except in special cases. Nevertheless, utilizing certain monotonicity properties, we show that they possess a unique equilibrium point; see Theorem 3.3. Moreover this equilibrium can be characterized as optimizing the sum (over mobiles) log throughputs.

The rest of the paper is as follows. In section 2 a generalized version of PFS is detailed. In section 3 we outline the key results. Section 4 gives numerical data. Only a brief outline of the ideas is given. Weaker conditions, applications to multitransmitter systems, and general data queues will be given elsewhere.

2 Proportional Fair Algorithm Outline

There are N mobiles transmitting data over a single (say, wireless) channel, and the possible rates of transmission of the individual users are randomly time varying. Time is divided into small scheduling intervals and mobiles are scheduled one at a time. If mobile i is selected in interval n , then it transmits $r_{i,n}$ units of data, where $\{r_{i,n}, n < \infty\}$ is a bounded (and usually correlated) random sequence, which might also be correlated among the i . They need only satisfy some mixing type condition, to be specified in the next section. It is assumed that each user has an infinite backlog of data, which has been the standard assumption in the literature to date. The end of scheduling interval n is called time n . Define the throughput up to time n for user i as .

$$\theta_{i,n} = \frac{1}{n} \sum_{l=1}^n r_{i,l} I_{i,l}, \quad (1)$$

where $I_{i,l} = 1$ if user i is chosen at time l and is zero otherwise. With the definition $\epsilon_n = 1/(n+1)$, (1) can be written in the recursive form (which defines Y_n)

$$\theta_{i,n+1} = \theta_{i,n} + \epsilon_n [I_{i,n+1} r_{i,n+1} - \theta_{i,n}] = \theta_{i,n} + \epsilon_n Y_{i,n}. \quad (2)$$

We will also treat the discounted alternative to (2), namely,

$$\theta_{i,n}^\epsilon = (1-\epsilon)^n \theta_{i,0}^\epsilon + \epsilon \sum_{l=1}^n (1-\epsilon)^{n-l} r_{i,l} I_{i,l}, \quad (3)$$

where ϵ is a small positive discount factor. This can be written in the recursive form (which defines Y_n^ϵ)

$$\theta_{i,n+1}^\epsilon = \theta_{i,n}^\epsilon + \epsilon [I_{i,n+1} r_{i,n+1} - \theta_{i,n}^\epsilon] = \theta_{i,n}^\epsilon + \epsilon Y_{i,n}^\epsilon. \quad (4)$$

The representations (2)-(4) allow arbitrary initial conditions which might reflect some past history.

The representation (2) allows the use of other values of ϵ_n . For example, they might go to zero more slowly than $1/n$, which provides for a weighing between those of (2) and (4). If the initial condition is zero (no past history), then it makes sense that the weights in (3) sum to unity. Then one would use the normalized form

$$\theta_{i,n}^\epsilon = \epsilon \sum_{l=1}^n (1-\epsilon)^{n-l} r_{i,l} I_{i,l} / [1 - (1-\epsilon)^n]. \quad (5)$$

Owing to the boundedness of the $r_{i,n}$, the solutions to (3) and (4) are bounded, provided that the initial conditions are confined to a compact set.

The original proportional-fair sharing algorithm chooses the user at time n which maximizes in

$$\arg \max \{r_{i,n+1} / \theta_{i,n}, i \leq N\} \quad (6)$$

or with $\theta_{i,n}^\epsilon$ if (4) is used.

When all of the current components $\theta_{i,n}, i \leq N$, are very small, there is little sense in (6), since in any practical sense, the current throughputs are all essentially zero and there is little motivation to distinguish between them. We modify the algorithm slightly as follows. Let $d_i, i \leq N$, be positive numbers, which can be as small as we wish. The chosen user at time n is that which maximizes in

$$\arg \max \{r_{i,n+1}/(d_i + \theta_{i,n}), i \leq N\} \quad (7)$$

In the event of ties, we randomize among the possibilities in order to resolve conflicts. The end results will be seen to be completely independent of how the conflicts are resolved. Define the vectors $\theta_n = \{\theta_{i,n}, i \leq N\}$ and $R_n = \{r_{i,n}, i \leq N\}$.

Definitions. The usual stochastic approximation asymptotic (or large time) analysis of the algorithms (2), (3) uses continuous time interpolations. For each n , define the *shifted* process $\theta^n(\cdot)$ (with components $\theta_i^n(\cdot), i \leq N$) by $\theta^n(0) = \theta_n$ and, for $l \geq 0$,

$$\theta^n(t) = \theta_{n+l} \text{ for } t \in \left[\sum_{k=n}^{n+l-1} \epsilon_k, \sum_{k=n}^{n+l} \epsilon_k \right), \quad (8)$$

where the empty sum is defined to be zero. Since the interpolated process $\theta^n(\cdot)$ starts at iterate n , the behavior of $\theta^n(\cdot)$ as $n \rightarrow \infty$ is that of θ_n as $n \rightarrow \infty$.

Define the interpolated process $\theta^\epsilon(\cdot)$ (with components $\theta_i^\epsilon(\cdot), i \leq N$) by $\theta^\epsilon(t) = \theta_n^\epsilon$ for $t \in [n\epsilon, n\epsilon + \epsilon)$.

3 Main Results

3.1 Assumptions

Assumption 3.2 is used only to assure that when a component θ_i is small there is a nonzero chance that user i will be chosen, no matter what the values of the other components of θ . It guarantees that $\bar{g}_i(\theta)$ is always positive when θ_i is small, hardly a restriction. The density assumption (3.3) and assumption (3.1) are satisfied under standard physical assumptions: for example, if the channel variations are due to Raleigh fading. The density condition is used only to show that the limit point is unique. Condition (10) is a very weak form of the law of large numbers, due to the use of the conditional expectation E_n . It basically says that the mean transmitted rate for user i on an interval $[n, n+m]$, conditioned on the data to n , converges to the ergodic average as m becomes large, hardly a restriction. If the conditional expectation of the transmitted rate at time l , given the data to time n is close to its stationary expectation for large $l-n$, then it holds. If the channel rate process is ergodic, then the condition holds even without the conditional expectation. So the combination of the conditional expectation and the division by m gives a very weak condition indeed.

Assumption 3.1 Let ξ_n denote the past: $\{R_l : l \leq n\}$. For each i, n , the function on \mathbb{R}_+^N defined by

$$g_{i,n}(\theta, \xi_n) = E_n r_{i,n+1} I_{\{r_{i,n+1}/(d_i+\theta_i) \geq r_{j,n+1}/(d_j+\theta_j), j \neq i\}}$$

is continuous in $\theta \in \mathbb{R}_+^N$. Here θ is considered fixed, and not random. Let $\delta > 0$ be arbitrary. Then in the set $\{\theta : \theta_i \geq \delta, i \leq N\}$, the continuity is uniform in n and in ξ_n .¹

Assumption 3.2 $\{R_n, n < \infty\}$ is stationary. Define the functions $\bar{g}_i(\cdot), i \leq N$, on \mathbb{R}_+^N by the stationary expectation:

$$\bar{g}_i(\theta) = E r_i I_{\{r_i/r_j \geq (d_i+\theta_i)/(d_j+\theta_j), j \neq i\}} \quad (9)$$

In (9), θ is considered fixed and not random. For $\theta \in \mathbb{R}_+^N$, The function $\bar{g}(\cdot)$ is continuous on \mathbb{R}_+^N . Also,

$$\lim_{m,n \rightarrow \infty} \frac{1}{m} \sum_{l=n}^{n+m-1} [E_n r_{i,l+1} I_{\{r_{i,l+1}/r_{j,l+1} \geq (d_i+\theta_i)/(d_j+\theta_j), j \neq i\}} - \bar{g}_i(\theta)] = 0 \quad (10)$$

in the sense of probability. There are small positive δ, δ_1 such that

$$P \{r_{i,n}/d_i \geq r_{j,n}/(d_j - \delta) + \delta_1, j \neq i\} > 0, \quad i \leq N. \quad (11)$$

Assumption 3.3 R_n is defined on some bounded set and has a bounded density.

It follows from assumption 3.3 that $\bar{g}(\cdot)$ is Lipschitz continuous.

3.2 Limiting ODE

The next theorem is by now a standard result in stochastic approximation. It basically says that the limit points of the algorithm (3), (7) are contained in those of the ODE (12).

Theorem 3.1 (This is [8, Theorems 2.2 and 2.3, Section 8.2].) Assume algorithm (2), 3.1 and 3.2. Then for any bounded set of initial conditions, any subsequence of $\theta^n(\cdot)$ has a further subsequence that converges weakly to the set of limit points of the solution of the ODE

$$\dot{\theta}_i = \bar{g}_i(\theta) - \theta_i, \quad i \leq N. \quad (12)$$

The same conclusion holds if the $\epsilon_n = 1/(n+1)$ in (2) is replaced by a sequence of positive numbers such that $\epsilon_n \rightarrow 0, \sum_n \epsilon_n = \infty$, and where ϵ_n doesn't vary too fast in the sense that for some sequence $\alpha_n \rightarrow \infty$

$$\lim_n \sup_{0 \leq l \leq \alpha_n} \left| \frac{\epsilon_{n+l}}{\epsilon_n} - 1 \right| = 0.$$

¹This is [8, condition (A2.3), Chapter 8]. The cited theorem holds if there is a delay in the conditioning; i.e., E_{n-m} is used, for any integer m . It can be weakened further but this is enough for typical applications where assumption 3.3 holds.

For algorithm (4), the same conclusion holds for the sequence $\theta^\epsilon(\epsilon q_\epsilon + \cdot)$ for any sequence of integers q_ϵ .

The ODE depends on the channel statistics only through the joint distribution of the current rates, and is thus independent of the fading rate. It has a number of important properties which stem from the definition of PFS. In particular the ODE satisfies the *Kamke* (or simply the *K-condition*). A function $f(\cdot)$ is said to satisfy the *K-condition* if for any x, y, i , with $x \leq y, x_i = y_i$, we have $f_i(x) \leq f_i(y)$. In our case, $f(\theta) = \bar{g}(\theta) - \theta$, and the condition holds. The condition says nothing more than the following, which is obvious for PDF: If θ_i is suddenly increased, then the other users are not less likely to be chosen. The *K-condition* implies the following monotonicity result. Its proof in [9, Proposition 1.1] assumes continuous differentiability of $f(\cdot)$. For our purposes, the main consequence of the *K-condition* is the following monotonicity theorem.

Theorem 3.2 [9, Proposition 1.1] *Let $f(\cdot)$ be Lipschitz continuous and assume the *K-condition*. If $x(0) \leq y(0)$ (resp., $<, \ll$, then $x(t|x(0)) \leq x(t|y(0))$ (resp., $<, \ll$).*

3.2.1 Two User example

Consider two independent users with received signal power determined by a stationary Rayleigh fading and with constant external noise. Suppose further that their rate declarations are proportional to the signal to noise ratio, with mean rates $1/\beta_i, i = 1, 2$ respectively. Then the ODE (12) becomes,

$$\begin{aligned} \dot{\theta}_1 &= \frac{1}{\beta_1} - \frac{\beta_1(d_1 + \theta_1)^2}{(\beta_1(d_1 + \theta_1) + \beta_2(d_2 + \theta_2))^2} - \theta_1 \\ \dot{\theta}_2 &= \frac{1}{\beta_2} - \frac{\beta_2(d_2 + \theta_2)^2}{(\beta_1(d_1 + \theta_1) + \beta_2(d_2 + \theta_2))^2} - \theta_2 \end{aligned} \quad (13)$$

3.3 Uniqueness of Limit Point and Characterization

The following result establishes that the throughputs converge to unique limiting values (taking a period of order $1/\epsilon$ slots). As we will see this fact has consequence for performance modeling of PFS, particularly in the case of long file transfers.

Theorem 3.3 *Assume algorithm (3) with assumptions 3.1, and 3.2. The limit point $(\bar{\theta})$ of (12) is unique, irrespective of the initial condition. So the processes $\theta^n(\cdot)$ and $\theta^\epsilon(\epsilon q_\epsilon + \cdot)$ converge to $\bar{\theta}$ as $n \rightarrow \infty$ (resp., as $\epsilon \rightarrow 0$ and $\epsilon q_\epsilon \rightarrow \infty$).*

The previous theorem shows that there is a unique asymptotically stable limit point $\bar{\theta}$ of the ODE and algorithm. We have not addressed the optimality of the algorithm at all. Intuitively it can be seen that PFS is a steepest ascent algorithm for a strictly concave utility function,

$$U(\theta) \doteq \sum_i \log(d_i + \theta_i). \quad (14)$$

\leq SNR	-12.5	-9.5	-8.5	-6.5	-5.7	-4.0
Rate	0.0	38.4	76.8	102.6	153.6	204.8
\leq SNR	-1.0	1.3	3.0	7.2	9.5	
Rate	307.2	614.4	921.6	1228.8	1843.2	

Table 1: Rate vs. SNR for 1% packet loss (taken from [1])

The problem is that the allowed directions of ascent depend heavily on θ . Hence there is no a priori guarantee of any type of maximization. However, it can still be shown that the rule maximizes the asymptotic value of the utility function $U(\cdot)$ as in

Theorem 3.4 *The rule (7) maximizes $\lim_n U(\theta_n)$ with respect to all non-anticipative policies.*

4 Performance Results

4.1 Transient Behavior

The graphs are from simulations based on Raleigh fading, and the relation between the current rates and signal to noise ratios is taken from Table 1, which comes from [1]. Our first results depict the advantages to be gained by taking advantage of the current values of the time varying rates. In Figure 1, one set of curves corresponds to the transient behavior for three mobiles using table 1 and mean SNRs, -12dB,-2dB,-8dB, respectively, using algorithm (4). There are two sets of curves: those with solid lines and (the higher ones) those with dotted lines. The solid lines depict the throughputs if the SNRs (and hence the rates) are assumed to be constant at the average values. $\epsilon = 0.0001$. Initially slots are offered only to mobile 2, with the other two mobile throughputs exponentially decaying. Also there are two “switching times”. At the first slots are equally divided between mobiles 2 and 3 (0, 1/2, 1/2) whereas at the second the slots are divided (1/3, 1/3, 1/3). (This behavior is generic for constant rates.) The second set of curves (dotted lines/filled symbol) are obtained for Rayleigh fading with fading rate 6 Hz and the same mean SNRs, and using PFS. As expected these curves show significant throughput gains from scheduling. Since the dependence on rates is roughly linear on SNR, it is expected the slots will be approximately evenly divided as the users SNR all have exponential distributions.

4.2 Comparison with Solution to ODE

Consider two users with received signal power determined by a stationary Rayleigh fading and with constant external noise. Suppose further that their rate declarations are proportional to the SNR, with mean rates $1/\beta_1, 1/\beta_2$ respectively. Then the ODE coincides with (13), once again using algorithm (4),(6). For two such users, and initial throughput 250.0, figure 2 shows a sample path for θ the

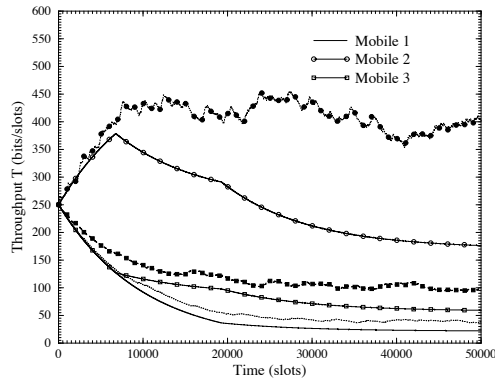


Figure 1: Time dependent behavior of Proportional Fair

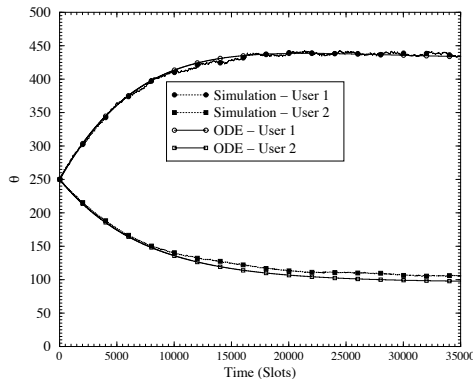


Figure 2: Sample path for θ and the Solution to the ODE

proportional fair throughput estimate and a numerical solution to the corresponding ODE. $\epsilon = 0.0001$ and rates are given via a Rayleigh fading simulator with $1/\beta_1 = 572$ bits/slot and $1/\beta_2 = 128$ bits/slot. The fading rates were taken as 60 Hz. In equilibrium the throughputs are, $\frac{3}{2} \cdot \frac{1}{2} \cdot 572 = 429$, $\frac{3}{2} \cdot \frac{1}{2} \cdot 128 = 96$. In general, from the equilibrium, for Rayleigh fading, the scheduling gains are given by $G(n) = \sum_{j=1}^n 1/j$.

The time constant for convergence is $1/\epsilon = 10,000$ slots and the results confirm convergence in a period of this order. The results also show the theoretical equilibrium being approached.

References

- [1] Paul Bender Peter Black et al. *CDMA/HDR A Bandwidth Efficient High Speed Wireless Data Service for Nomadic Users* IEEE Communications

Magazine 2000.

- [2] M. Benaïm and M.W. Hirsch. Learning processes, mixed equilibria and dynamical systems arising from repeated games. 1997.
- [3] M. Benaïm and M.W. Hirsch. Stochastic approximation algorithms with constant step size whose average is cooperative. *Ann. Appl. Prob.*, 9:216–241, 1999.
- [4] M.W. Hirsch. Systems of differential equations that are competitive and cooperative: Convergence almost everywhere. *SIAM J. Math. Anal.*, 16:423–439, 1985.
- [5] F.P.Kelly *Reversibility and Stochastic Networks* John Wiley, London, 1979.
- [6] F.P.Kelly, A.K. Maulloo and D.K.H. Tan. *Rate control in communication networks: shadow prices, proportional fairness and stability* *Journal of the Operational Research Society* 1998.
- [7] H.J. Kushner and P.A. Whiting. Asymptotic properties of proportional-fair sharing algorithms under general conditions. In preparation: Brown University, Applied Math. LCDS Report., 2002.
- [8] H.J. Kushner and G. Yin. *Stochastic Approximation Algorithms and Applications*. Springer-Verlag, Berlin and New York, 1997.
- [9] H.L. Smith. *Monotone Dynamical Systems: An Introduction to Competitive and Cooperative Systems*, *AMS Math. Surveys and Monographs*, Vol. 41. American Mathematical Society, Providence RI, 1995.
- [10] Matthew Andrews, Krishnan Kumaran, Kavita Ramanan, Alexandar Stolyar, P Whiting. *Providing Quality of Service over a Shared Wireless Link*. *IEEE Personal Communications*, January 2001.