

# Fair Resource Allocation with Varying Time Constraints

Sreenath Ramanath  
Lekha Wireless Solutions  
Bangalore, India

Eitan Altman  
INRIA  
Sophia Antipolis, France

**Abstract**—In this paper we discuss fair resource allocation among users with varying time constraints. Often, resource allocation problems study short-term (instantaneous) or long term (average) fairness where the resource allocation is over a given (same) period of time. What if we want to address fair sharing of resources when the users are interested in two different time periods? In this context, we study fair resource allocation among users with varying time constraints. We formulate this as a Markov Decision Process (MDP) for a two user case and provide a Dynamic Program (DP) solution. Simulation results in an LTE framework are provided to support the theoretical claims.

## I. INTRODUCTION

Resource allocation algorithms often try to achieve fairness and efficiency. In particular, Opportunistic scheduling algorithms in both downlink (base station to mobile) and uplink (mobile to base station) of cellular networks achieve higher throughput by giving preference to mobiles with better relative channel conditions. Preference according to relative channel conditions rather than absolute channel conditions mean that the radio conditions of each mobile are normalized by the averaged conditions of that channel until then, and mobiles with best normalized radio conditions are selected for transmission; this guarantees fairness.

Averaging the radio conditions is done in practice using some low pass filter. Thus averaging is done over some effective period  $T$ . A longer  $T$  achieves a larger opportunistic gain at the cost of longer starvation periods. In other words, efficiency is obtained at the cost of being more unfair over a short time scale. Whereas elastic traffic may prefer to be insensitive to short time scale unfairness, interactive real time applications may need averaging over shorter time. Hence each user desires fairness over a time scale that suits his need.

In our work, we want to address the problem of fairness in resource allocation where each user desires fair resource allocation over a time scale of his choice in a cellular network (See for e.g., Fig. 1). Our work is based on the heterogeneous timescale approach for fair resource allocation in wireless networks [1]. The notion of  $\alpha$ -fairness was introduced in [2]. This, as well as other fairness notions can be defined through a set of axioms, see [3]. For a rich literature on concepts of fairness and some interesting applications, the reader is referred to [4], [5], [6], [7], [9], [8] and the references there in.

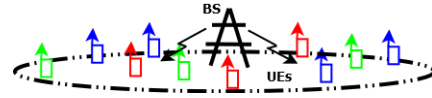


Fig. 1. Users in a Cellular Network with varying time constraints

## II. SYSTEM MODEL

In the previous section, we have addressed heterogeneous time in fair assignment by separating to different time scales - an instantaneous one and a time average one. By doing so we extended the standard definition of fairness which has not included any time consideration.

### A. Fairness among applications requiring averaging over different time periods

In this section, we refine the fairness differentiation according to time scales, and consider resource allocation problems in which an application needs neither instantaneous fairness nor average fairness (over the whole session duration) but a fairness over a well defined time period. We know that the larger the period over which fairness is defined, the better the performance is; this is due to the opportunistic gain.

As an example, consider two sources that have radio conditions allowing to transmit at the following rates during 4 slots:

$$X^1 = (10, 10, 1, 1); X^2 = (1, 1, 10, 10)$$

Assume fairness is required for the throughput averaged over the 4 slots. Then each user can receive 20 units by allocating the slots as (1,1,2,2). If we need to allocate fairly over intervals of 2 slots then the fair assignment allocates 11 units to each user. We therefore see that longer periods for averaging can result in better performance.

What happens if one user wishes allocation to be over the whole 4 slots and the other over periods of 2 slots? We shall study below such problems.

To motivate the definition of fairness over a given time period, consider a voice call competing with a large file transfer. For the file transfer, it is obviously the average transfer time that matters, or in other words, the throughput averaged over the whole transfer duration. Voice calls usually have a playout buffer that can absorb fluctuation in throughput. Thus voice call would benefit by requesting the fair assignment to

be over time periods that are of the order needed to empty or to fill their playout buffer.

We meet time periods related to averaging in actual implementation of fair assignment. Indeed, actual fair scheduling in wireless networks, and in particular the proportional fairness which has been implemented in many wireless standards (viz. HSDPA, LTE, etc.) attempt to obtain fairness of the time average assignment. Implementations achieve the time average through some sort of low pass filter in which the current average estimation  $\hat{x}(t+1)$  is updated by taking the weighted average between the current assignment  $y(t)$  and the previous estimation of the average  $\hat{x}(t)$ :

$$\hat{x}(t+1) = \epsilon y(t) + (1 - \epsilon)x(t)$$

This means that the averaging is done in practice over an interval of the order of  $1/\epsilon$ .

To define fair assignment of resources having different averaging periods, we shall need as input

- The time period  $T_\ell$  for each user  $\ell$ .
- The utility for assigning  $k_\ell$  units during  $T_\ell$  to user  $\ell$ .

### B. Formulation as a Markov Decision Process

- Call  $m_\ell$  consecutive slots that arrive from user  $\ell$  a block.
- Utility of user  $\ell$ ,  $U_\ell$ : the user needs to get  $k_\ell$  successful time slots out of  $m_\ell$  consecutive ones.
  - If it does then its utility corresponding to the successful transmission in the time period  $m_\ell$  is 1.
- Let  $P(r_\ell)$  denote the probability that the radio channel of user  $\ell$  corresponds to transmitting at a rate  $r$ .
  - We assume that  $r$  takes finitely many values  $r \in \{r(1) < r(2) < \dots < r(R)\}$ .

Assume that user  $\ell$  wishes to obtain fairness over a time period of  $m_\ell$  units,  $\ell = 1, 2$ . Take  $T$  to be the smallest number which is an integer multiple of  $m_\ell$  for both  $\ell = 1, 2$ . Introduce the following MDP:

- **State:**  $s = (\zeta_\ell, r_\ell, \rho_\ell; \ell = 1, 2)$ .
  - $r_\ell$  is the rate at which user  $\ell$  would transmit if scheduled at present.
  - $\rho_\ell$  the total rate that user  $\ell$  has already received in the current period of length  $m_\ell$ .
  - $\zeta_\ell$  is the total utility obtained by user  $\ell$  over previous blocks.
- **Action:**  $a \in \{1, 2\}$  The action  $\ell$  means to allocate a slot to user  $\ell$ .
- **Transition Probabilities:**  $P_{s,a,s'}(t)$  is given by

$$\prod_{\ell=1,2} P(r'_\ell) \times 1_{\{\rho'_\ell = \rho_\ell 1_{\{t \text{ not ISM of } m_\ell\}} + r_\ell 1_{\{a=\ell\}}\}} \times 1_{\{\zeta'_\ell = \zeta_\ell + U_\ell (\rho_\ell + r_\ell 1_{\{a=\ell\}})\}} 1_{\{t \text{ ISM of } m_\ell\}}$$

where,  $s = (\zeta_1, \zeta_2, \rho_1, \rho_2, r_1, r_2)$   
 $s' = (\zeta'_1, \zeta'_2, \rho'_1, \rho'_2, r'_1, r'_2)$  and  
ISM - Integer sub-multiple

Note: we can consider the more general case of Markov dependence between the rates at different times instead of the i.i.d assumption. More precisely, assume that the probability that the radio channel of user  $\ell$  has a potential rate of  $r'_\ell$  given that it had  $r_\ell$  in the last slot is given by  $P(r_\ell \rightarrow r'_\ell)$ . Then the only change in the above will be in the transition probabilities where  $P(r_\ell \rightarrow r'_\ell)$  will replace  $P(r'_\ell)$ ,  $\ell = 1, 2$ .

### C. Dynamic Programming

We formulate a dynamic program to solve the MDP. Policy  $\pi$  gives an action for each state at each time. We want to find the optimal policy  $\pi^*$ , which maximizes the expected sum utility. Let  $V_t(s)$  be the expected sum utility accumulated when starting from state  $s$  and acting optimally for a horizon of  $t$  steps.

The dynamic programming is now given by:

$$V_{t+1}(s) = \max_{a=1,2} \sum_{s'} P_{s,a,s'}(t) V_t(s') \quad (1)$$

with

$$V_0(s) = \sum_{\ell=1,2} \frac{\zeta(\ell)^{1-\alpha}}{1-\alpha} \quad (2)$$

Note that the argument of the summation in equation (2) is the  $\alpha$ -fair utility of user  $\ell$ .

## III. SIMULATION MODEL

We use an LTE framework to simulate fair resource allocation with varying time constraints. An LTE frame (See Fig. 2) is 10 msec long and constitutes 10 TTIs (Transmission Time Intervals), each of 1 msec. Each TTI is further divided into two slots of 0.5 msec each. Every slot is a frequency-time grid comprising the basic units of resource allocation known as RBs (resource blocks). Information for a given user is coded, modulated (QPSK, 16-QAM or 64-QAM) and mapped on to the REs (the smallest information carrying unit) of an RB. Any user can be allocated plurality of resource blocks with the constraint limited by the bandwidth of operation. LTE supports 1.4, 3, 5, 10, 15 and 20 MHz bandwidth.

For our simulation, we consider a 10 MHz LTE system, which supports 50 RBs in a slot or 100 RBs in each TTI. A part of these RBs (about 10-15%) are used for control channels and pilots (intended for channel estimation and synchronization purposes) and rest is available for data channels. So every TTI, the scheduler after ear marking few RBs for control information, can distribute the available RBs to the connected users based on user specific demand (QoS, Traffic type, CSI, etc.). In our simulations, we quantify the CSI with a channel quality indicator (CQI), which helps in choosing the most optimal modulation and coding scheme for the intended transmission.

The simulation is carried out over multiple LTE frames and a snap shot of resource allocation and throughput achieved is captured in Fig 3 and 4.

In the first part of the simulation, we use a proportionally-fair scheduler [5] based on the channel state information (CSI)

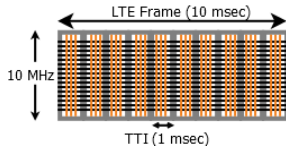


Fig. 2. Time-Frequency Resource grid of an LTE frame

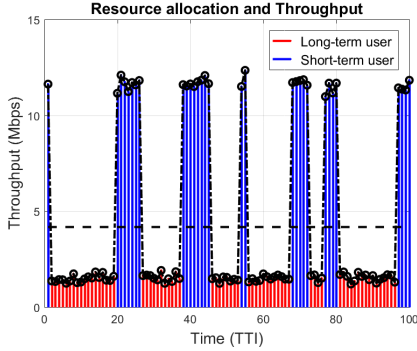


Fig. 3. Resource allocation and Throughput with Proportional-fair Scheduling

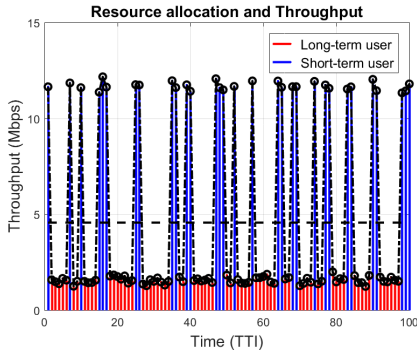


Fig. 4. Resource allocation and Throughput with Proportional-fair Scheduling under varying time constraints

of each user and allocates. The resource allocation and user throughput is captured in Fig. 3.

In the second part of the simulation, our proposed scheduler (fairness under varying time constraints), with  $\alpha$  set to 1 for proportional fairness [2] is used in conjunction with the policy derived to satisfy long-term and short-term fairness for two classes of users. The scheduling constraint for the short term user is set to 10 TTI (10 msec) and for the long term user as 100 TTI (100 msec).

The scheduling and throughput information is captured in Fig. 4. Comparing the two schedulers, both under proportional fairness, but ours tuned for varying time fairness, we observe that the short term user is served more uniformly every LTE frame to satisfy the demand, while the long term user is served at a steady rate. In contrast, the proportional fair scheduler does not guarantee fairness over varying time scales, specifically for the short term user. The proposed scheduler achieves fairness under varying time constraints, without compromising on the average system throughput (dashed horizontal black marker in Fig 3 and 4).

#### IV. CONCLUSION AND PERSPECTIVES

In this study we discuss fair resource allocation when users demand fairness over different time scales. We formulate a Markov Decision Process and provide a Dynamic programming (DP) solution. The underlying problem formulation is for a two user case. Characterization of the DP solution, simulation results and problem formulation for the Multi-user case with varying time constraints is being addressed.

#### REFERENCES

- [1] E. Altman, K. Avrachenkov and S. Ramanath, "Multiscale fairness and its application to resource allocation in wireless networks", Vol 35, Issue 7, Apr 2012, Elsevier Computer Communications
- [2] J. Mo and J. Walrand, "Fair End-to-end Window-based Congestion Control", IEEE/ACM Transactions on Networking, Vol. 8, No. 5, pp. 556-567, 2000.
- [3] T. Lan, D. Kao, M. Chiang and A. Sabharwal, "An Axiomatic Theory of Fairness in Network Resource Allocation", In Proceedings of IEEE INFOCOM, 2010, San Diego, USA.
- [4] R. Jain, D. M. Chiu and W. R. Hawe, "A Quantitative Measure of Fairness and Discrimination for Resource Allocation in Shared Computer Systems", Technical Report TR 301, Digital Equipment Corporation, 1984
- [5] F.P. Kelly, A.K. Maulloo and D.K.H. Tan, "Rate control for communication networks: shadow prices, proportional fairness and stability", Journal of the Operational Research Society, Vol. 49, No. 3, pp. 237-252, 1998.
- [6] M. J. Neely, E. Modiano and C. Li, "Fairness and Optimal Stochastic Control for Heterogeneous Networks", In Proceedings of IEEE INFOCOM, 2010, San Diego, USA.
- [7] A. Eryilmaz, R. Srikant, "Fair resource allocation in wireless networks using queue-length-based scheduling and congestion control", IEEE/ACM Transactions on Networking, vol. 15, pp. 1333-1344, 2007
- [8] E. Altman, K. Avrachenkov and A. Garnaev, "Generalized Alpha-Fair Resource Allocation in Wireless Networks", In Proceedings of 47th IEEE Conference on Decision and Control, 2008, Cancun, Mexico.
- [9] C. J. Wong, S. Sen, T. Lan, M. Chiang, "Multiresource Allocation: Fairness vs Efficiency Tradeoffs in a Unifying Framework", IEEE/ACM Transactions on Networking, vol. 21, pp. 1785-1798, 2013