Heuristic approach for forecast scheduling

Hind Zaaraoui*, Zwi Altman*, Sana Ben Jemaa*, Eitan Altman[†] and Tania Jimenez[‡]

*Orange Labs

38/40 rue du Général Leclerc, 92794 Issy-les-Moulineaux

Email: {hind.zaaraoui, zwi.altman, sana.benjemmaa}@orange.com

[†]INRIA Sophia Antipolis and LINCS, France

Email: Eitan.Altman@sophia.inria.fr

[‡]Avignon University

Email: tania.jimenez@univ-avignon.fr

Abstract—Forecast Scheduling (FS) is a scheduling concept that utilizes rate prediction along the users' trajectories in order to optimize the scheduler allocation. The rate prediction is based on Signal to Interference plus Noise Ratio (SINR) or rate maps provided by a Radio Environment Map (REM). The FS has been formulated as a convex optimization problem namely the maximization of an α -fair utility function of the cumulated rates of the users along their trajectories [1]. This paper proposes a fast heuristic for the FS problem based on two FS users' scheduling. Furthermore, it is shown that in the case of two users, the FS problem can be solved analytically, making the heuristic computationally very efficient. Numerical results illustrate the throughput gain brought about by the scheduling solution.^a

Index Terms—Forecast scheduler, alpha-fair, high mobility, Radio Environment Maps, geo-localized measurements

I. INTRODUCTION

The concept of FS has been introduced in [1], where it is shown that the knowledge of the present and future rates along the users' trajectories can be exploited by the scheduler in order to significantly improve the average user throughput. The throughput gain is achieved by exploiting long term time and spatial diversity along the users' trajectory. The FS allocation is posed as a convex optimization problem that can be solved using fast convex optimization solvers. The randomness of the traffic, i.e. arrival and departure of communications on the one hand and randomness in the user trajectories on the other hand can be incorporated into the FS solution [2].

The SINR or rate along the mobile users' trajectories can be provided using a REM. The generation of REMs from Geo-Localized Measurements (GLM) has been the topic of recent research activity in both industry and

academia and is described in more details in Section II. While abundant work on REM construction has been published, little work is available on how to exploit GLM and REMs to better manage the network.

The reasons for the growing interest in exploiting GLM are the following: First, it allows network operators to better assess the actual Quality of Service (QoS) of mobile users at any location; Second, it provides significant levers to optimize the network performance; and third, it allows personalized, user centric type of optimization. GLM can feed new Radio Resource Management (RRM) algorithms, and in general, a management entity to (self-) configure, optimize and troubleshoot the network. The use of GLM corresponds to a general trend in 5G networks, namely the exploitation of data from different sources in order to improve the network operation.

The purpose of this paper is to introduce a fast heuristic for the problem of calculating the FS allocation. Such a heuristic can simplify the implementation of the FS in real equipment. The contributions of this paper are the following:

- Provides an analytical solution for the FS allocation problem for the case of two mobile users;
- Develops a computationally fast FS allocation heuristic based on the solution for two mobile users.

The paper is organized as follows: Section II provides an overview of REMs and illustrates how the FS can exploit the REM to achieve throughput gain. Section III recalls the basic FS formulation and provides the main results for a closed form allocation rule for the case of two users. Section IV presents a two users based FS heuristic. Numerical results are described in Section V followed by concluding remarks in Section VI.

II. RADIO ENVIRONMENT MAPS

With the availability of Global Positioning System (GPS) information in most mobile terminals, it has

^aThis work has been partially carried out in the framework of IDEFIX project, funded by the ANR under the contract number ANR-13-INFR-0006.

become possible to retrieve radio measurements together with their corresponding location. GLMs can be reported to the network using User Equipments (UEs) with chipsets that implement the Minimization of Drive Testing (MDT) feature standardized by the 3rd Generation Partnership Project (3GPP) [3]. GLM comprises a rich source of information that allows network operators to optimize the network with finer granularity than the classical cell level, and to go one step further towards personalized optimization of the end-user experience.

As the measurements are performed by UEs, the operator does not have the control on the positions where these measurements are performed. Hence the first challenge has been to complement the information provided by GLM namely predicting the considered quantity in the entire area of interest and to build the complete REM.

The concept of REM was first introduced in cognitive radio and TV whitespace database systems [4], [5]. In mobile networks, the main research has focused on building reliable maps that could be trusted by radio engineers and that could be integrated in automated optimization processes and tools.

Several works ([6], [7], [8]) have considered the geostatistical interpolation technique known as Kriging [9] to build REMs. This technique exploits the spatial correlation of the signal propagation to provide higher prediction accuracy than classical propagation models. The accuracy of the prediction can be enhanced by increasing the density of the measurement samples. However, the computational complexity of the algorithm increases rapidly with the number of measurement points n , namely as $O(n^3)$. The reduction of the computational complexity has been among the main research challenges. The Fixed Rank Kriging (FRK) [10] solution brings the computational complexity of the REM construction to a linear level with respect to n , while maintaining high quality. The latter can be measured in terms of the root-mean square error (RMSE) of the interpolated GLM with respect to a reference test set of GLM for example. The coverage map is obtained using a ray tracing propagation model.

Figure 1(a) shows GLM of the received signal strength over 10 percent of the meshes in a cell area, and the predicted coverage area obtained using the FRK algorithm in Figure 1(b). In this example, the FRK is able to build the entire map with a root-mean square error not exceeding 3 dB.

Once the received signal strength coverage map is built with a satisfactory accuracy, it can serve as input to construct other radio measurement maps such as the SINR map. Figure 2(a) illustrates an SINR map with two trajectories drawn in blue and black. In Figure 2(b)

Fig. 1. Coverage map prediction using FRK.

one can see the reachable bitrates that two vehicles will encounter while moving along the two trajectories using the Shannon formula. The significant spatial diversity in SINR and in the corresponding rate along the trajectories motivates the development of computationally efficient FSs.

Fig. 2. From SINR map to reachable bitrates.

III. FORECAST SCHEDULING MODEL AND ANALYTICAL RESULTS

In this Section we first briefly recall the basic formulation of the FS as presented in [1]. Then, using Karush-Kuhn-Tucker (KKT) conditions, we provide the main results for the closed form solution for $n = 2$ users. Due to space limitation, the complete formulation is given in a supplementary document [11].

A. Basic formulation

Consider a macro-cell (Base Station (BS)) surrounded by interfering BSs. We suppose that the SINR of the users in mobility at any location is provided to the BS by a REM. Consider n full buffer users moving at a constant speed during a time interval T - the scheduling period, over which n is considered constant. Suppose that time is in a discrete space: $t \in \{1, 2, \dots, T\} = [1, T]$ and let i denote the user number, $i \in \{1, 2, \dots, n\} = [1, n]$. We suppose that during the scheduling duration there are no arrivals or departures of users (it is noted that this assumption can be relaxed [2]).

A scheduling period T (typically of the order of seconds) is divided into scheduling time slots e.g. of 1ms, during which the bandwidth is shared among the scheduled users. In practice, the time resolution of the FS is longer than the system time slot (e.g. of 1ms), and is referred to hereafter as time slot (see Section V). Let $a_i(t)$ denote the bandwidth proportion allocated to the user i at time t , $a_i(t) \in [0, 1]$ according to the scheduling strategy, where $\forall t$, $\sum_{i=1}^n a_i(t) = 1$, and W is the total bandwidth. Using the Shannon equation, the throughput as a function ϕ of the SINR of user i reads

$$a_i(t)\phi(\text{SINR}_i(t)) = a_i(t)W \log_2(1 + \text{SINR}_i(t)). \quad (1)$$

Denote by S_i^t the predicted SINR (i.e. the one provided by the REM). The FS allocation policy is defined by the following optimization problem, with $\alpha \neq 1$:

$$\begin{aligned} \text{maximize : } f(a) &= \sum_{i=1}^n \frac{(\sum_{t=1}^T a_i(t)\phi(S_i^t))^{1-\alpha}}{1-\alpha} \\ \text{subject to : } &\forall i, \forall t, a_i(t) \geq 0 \\ &\forall t, \sum_{i=1}^n a_i(t) = 1. \end{aligned} \quad (2)$$

α is a positive fixed parameter of the optimization problem. For $\alpha \rightarrow 1$, the optimization problem with the same constraints reads:

$$\text{maximize : } f(a) = \sum_{i=1}^n \log(\sum_{t=1}^T a_i(t)\phi(S_i^t)). \quad (3)$$

Both equations (2) and (3) have concave functions f for $\alpha \geq 0$, and can be solved using convex optimization (e.g. CVX solver [12], see more details in Section V). The size of the optimization problem is defined by the number of unknown variables, namely $n \times T$ (with T being the number of time slots).

The interpretation of (2) and (3) is the following: resources are shared fairly among the users according to the data-rate variation along their trajectories. For example, if a user has a large enough coverage hole in his future trajectory, the FS may allocate to this user as much data as possible before reaching the coverage hole so as to remain fair with respect to the other users.

B. KKT conditions

We first derive the KKT conditions for the problem (2). The optimization problem (2) has a nonlinear objective function f with regular equality and inequality conditions i.e. differentiable constraint functions. If the objective and constraint functions (2) are continuously differentiable at any $a = (a_1(1), \dots, a_1(T), a_2(1), \dots, a_n(T)) \in$

$\mathbb{R}^{nT}$, then there exist multipliers $\lambda_{k,j}$ and ν_j , where $k \in [1, n]$ and $j \in [1, T]$, called KKT multipliers ([13], Chap.5) with the following Lagrangian function:

$$L(a, \nu, \lambda) = f(a) + \sum_{j=1}^T \nu_j (\sum_{k=1}^n a_k(j) - 1) + \sum_{k,j} \lambda_{k,j} a_k(j), \quad (4)$$

where $\lambda_{k,j} \geq 0$.

We define the Lagrange dual function as the maximum value of L over a . Let a^* maximize the Lagrangian function (4) for the optimal multipliers $\lambda_{k,j}^*$ and ν_j^* , where $k \in [1, n]$ and $j \in [1, T]$, and therefore its gradient is null at this point:

$$\nabla L(a^*, \nu^*, \lambda^*) = 0. \quad (5)$$

From (5) one obtains for all $i \in [1, n]$ and $t \in [1, T]$ the corresponding KKT conditions:

$$\phi(S_i^t) (\sum_{j=1}^T a_i^*(j) \phi(S_i^j))^{-\alpha} + \nu_t^* + \lambda_{i,t}^* = 0, \quad (6)$$

$$a_i^*(t) \geq 0, \quad (7)$$

$$\sum_{k=1}^n a_k^*(t) = 1, \quad (8)$$

$$\lambda_{i,t}^* a_i^*(t) = 0 \quad (9)$$

$$\lambda_{i,t}^* \geq 0. \quad (10)$$

We note from (6) that for any time t , for all users i and w with $w \neq i$

$$\begin{aligned} &\phi(S_i^t) (\sum_{j=1}^T a_i^*(j) \phi(S_i^j))^{-\alpha} + \lambda_{i,t}^* \\ &= \phi(S_w^t) (\sum_{j=1}^T a_w^*(j) \phi(S_w^j))^{-\alpha} + \lambda_{w,t}^* \end{aligned} \quad (11)$$

and also from (6), for all user i , we have for any time t and u :

$$\frac{\lambda_{i,t}^* + \nu_t^*}{\phi(S_i^t)} = \frac{\lambda_{i,u}^* + \nu_u^*}{\phi(S_i^u)} \quad (12)$$

as ν_t^* does not depend on the users, and $\sum_{j=1}^T a_i^*(j) \phi(S_i^j)$ does not depend on the time. Equality (11) gives explicitly the resource balancing among users at each time relatively to α in the sense of equalizing the two expressions for each user.

C. Analytical solution for $n = 2$ users

In the calculations that follow, we rewrite $a = a^*$, $\lambda = \lambda^*$ and $\nu = \nu^*$. We assume two users in the cell, i.e. $n = 2$ in problem (2).

KKT analysis shows that there exists at most one time slot $K \in [1, T]$ for which two users can be scheduled simultaneously [11]. For all other times in $[1, T]$ one user is scheduled at a time. We formulate hereafter the more frequent case where there is no $K \in [1, T]$ for which two users scheduled at the same time. The complete analysis is available in [11].

We suppose in this section that at time $u = 1$ user 1 is scheduled and therefore $\lambda_{1,1} = 0$. We need to verify this assumption, namely that user 1 is indeed selected at time $u = 1$ (see Theorem 2). For all time t , the sign of $\lambda_{1,t} - \lambda_{2,t}$ determines the scheduling decision: if it is positive then user 2 is scheduled at time t ($\lambda_{1,t} \geq \lambda_{2,t} \iff a_2(t) \geq a_1(t)$ by equations (9-10)). One of the lambdas must be null and the other one positive for the case of two users.

Using equation (12), as $\lambda_{1,1} = 0$, we have:

$$\lambda_{1,t} - \lambda_{2,t} = \nu_1 \psi_1^{t/1} - \nu_1 \psi_2^{t/1} - \lambda_{2,1} \psi_2^{t/1} \quad (13)$$

where $\psi_i^{t/u} = \frac{\phi(S_i^t)}{\phi(S_i^u)}$. We have then:

$$\lambda_{1,t} - \lambda_{2,t} = \nu_1 (\psi_1^{t/1} - \psi_2^{t/1}) - \lambda_{2,1} \psi_2^{t/1} \quad (14)$$

As $\lambda_{2,1} \psi_2^{t/1} > 0$ we divide equation (14) by $\lambda_{2,1} \psi_2^{t/1}$ and hence the sign studied is the same as the sign of $\frac{\psi_1^{t/1} - \psi_2^{t/1}}{-\frac{\lambda_{2,1}}{\nu_1}} - 1$. Note that from (6), ν_1 is negative.

If $\psi_2^{t/1} < \psi_1^{t/1}$ then $\lambda_{1,t} - \lambda_{2,t} < 0$ hence user 1 will be scheduled at time t . We cannot say more using this method.

Theorem 1. *If user 1 is supposed scheduled at time $u = 1$, then:*
 $\psi_2^{t/1} < \psi_1^{t/1} \implies \text{user 1 is scheduled at time } t$.

Denote by A_1^u and A_2^u the following sets:

$$\begin{aligned} A_1^u &= \{t \neq u, \psi_1^{t/u} > \psi_2^{t/u}\} \cup \{u\} \\ A_2^u &= \{t \neq u, \psi_1^{t/u} < \psi_2^{t/u}\} \end{aligned}$$

For $u = 1$ and using Theorem 1 i.e. if $t \in A_1^u$ then we have that user 1 is surely scheduled at time $t > 1$ with the assumption that the user is scheduled at time 1. We have the following result:

Theorem 2. *User 1 is scheduled at time u if and only if:*

$$\left(\sum_{A_2^u} \phi(S_2^j) \right)^\alpha > \max_{k \in A_1^u} \frac{\phi(S_2^k)}{\phi(S_1^k)} \left(\sum_{A_1^u} \phi(S_1^j) \right)^\alpha \quad (15)$$

Equivalent scheduling rule can be written for user 2. The proof of the theorem is given in [11].

IV. HEURISTIC SOLUTION FOR FORECAST SCHEDULING

In this section, we propose a heuristic FS solution for an arbitrary number of users n which relies on FS with $n = 2$ users.

Consider an iterative algorithm for scheduling user(s) at time t with two main steps: (i) choose two users, i_1 and i_2 with the highest α -fair utility given by equations (16) and (17) where $R_j(t)$ is the mean past data rate (calculated over some time window) of the user j until time t ; (ii) Select the best user (or both users) according to the FS rule for $n = 2$ for the time interval $[1, T]$. It is recalled that FS for $n = 2$ can be computed using closed form formula or using a fast convex optimization solver. The algorithm is depicted in Table I.

$$i_1(t) = \underset{j \in \text{users}}{\operatorname{argmax}} \frac{\phi(S_j^t)}{R_j(t)^\alpha} \quad (16)$$

$$i_2(t) = \underset{j \in \{\text{users}\} - \{i_1(t)\}}{\operatorname{argmax}} \frac{\phi(S_j^t)}{R_j(t)^\alpha}, \quad (17)$$

Fig. 3. Scheduling using the best two α -fair heuristic.

The user selection algorithm in Table I is not as optimal as the direct FS technique since it considers only the future SINR trajectories of the two best α -fair users. Interestingly, this approach provides significant gains with respect to Round Robin (RR) scheduler. It is recalled that at high mobility, the Proportional Fair (PF) scheduling converges to RR due to the short coherence time of the channel.

It is noted that alternatively and with a higher complexity, Step 2 in Table I can be performed using the FS for the two users selected in step 1 using a convex optimization solver.

TABLE I
TWO BEST USERS ALGORITHM (TBUA)

Algorithm of approximation with two best α -fair users	
Step 1	-Select the two best users i_1 and i_2 according to (16) and (17) at time t
Step 2	-Apply KKT resolution for the case of two users for i_1 and i_2 . Verify if time K exists (using (19) in [11]). If it exists, use allocation (16-18) in [11]. Otherwise use (15), and select the user to schedule at time t accordingly (using the condition in Theorem 1).
Step 4	-Update the mean past data rate for all users utilizing the received data rate at time t
Step 5	-if $t+1 \leq T$ proceed to time $t+1$, T being the scheduling period

Fig. 4. A road crossing a cell and the coverage area of a VSC.

V. NUMERICAL RESULTS

This Section presents numerical results for the FS heuristic obtained using a Long Term Evolution (LTE) network simulator written in Matlab. We consider a scenario with 10 vehicular users moving with a constant speed of 50km/h along a road crossing the cell. Close to the cell edge, the road traverses a coverage area of a Virtual Small Cell (VSC). It is recalled that the VSC is a remotely created small cell using a Large Scale Antenna System (LSAS) [14] which is used to enhance the spatial SINR diversity along the trajectory (see Figure 4).

The spatial resolution of the REM is of 1m (which is feasible thanks to the interpolation of the GLM). The corresponding time interval at 50km/h is 70ms over which the SINR is considered constant. This time interval defines the time resolution of the FS during which a fixed allocation is applied, i.e. the same users are scheduled at time intervals defined by the technology (e.g. 1ms for LTE). The SINR experienced by the users along the road is presented in Figure 5. The first (stronger) and second maxima correspond to the transmissions at the vicinity of the VSC and the macro BS respectively. Simulation parameters are depicted in Table II.

In the following, we compare the TBUA solution with the (standard) FS solution of problem 3. The FS solution is solved using CVX (details below). The TBUA is solved using the heuristic solution described in Section IV with the analytical FS of Section III-C.

Fig. 5. SINR for different users along the road.

TABLE II
SIMULATION PARAMETERS

Network parameters	
Number of macro-cell BSs	1
Number of interfering BSs	6
Macro-cell layout	hexagonal omni sectors
Intersite distance	500 m
Bandwidth	20 MHz
Channel characteristics	
Thermal noise	-174 dBm/Hz
Macro Path loss (d in km)	$128.1 + 37.6 \log_{10}(d)$ dB
Mobility traffic characteristics	
User speed	50 km/h
Number of users	10
File size σ	full buffer (∞)
One iteration = Scheduling delay	70 ms

We assume that the interfering neighboring cells have 10 percent of load (details on the interference model can be found in [1]). The interference conditions when using the REM will differ from those used to create the REM. We denote this as *interference errors* in the sequel. To show the robustness of the FS with respect to interference errors, we consider two cases: (i) a REM constructed with fully (100 percent) loaded interfering cells, which is denoted hereafter as "*with interference errors*" and (ii) a hypothetical REM that provides the exact SINR, namely 10 percent of loaded interfering cells, and is denoted as "*without interfering errors*".

Reference [1] explains that the scheduling decisions of the FS can be impacted to a certain extent by the interference errors, and the impact of these errors decreases with the increase in the scheduling periodicity. Moreover, the impact of the interference errors on the Mean User Throughput (MUT) is negligible. This behavior is explained by the high spatial dynamicity of SINR variations, the trends of which are well captured by the REM in spite of the interference errors. We consider the RR as a baseline for the different approaches. The rationale is that at high mobility considered here, the coherence time is too short to exploit fast fading in order to achieve opportunistic scheduling gain.

The objective functions of the optimization problems (2) and (3) are convex and can be solved using a convex

optimization solver. The CVX Matlab library has been used (see [12] and [15]). The CVX solver verifies the convexity of the problem and solves it using *SDPT3* or *SeDuMi*. *SDPT3* implements the infeasible path-following algorithms for solving conic programming problems whose constraint cone is a product of semi-definite cones. It uses a predictor-corrector primal-dual path-following method, with different types of search direction. *SeDuMi* is a linear/quadratic/semi-definite solver.

The MUT time evolution for the 10 users and for the three different FS solutions with 2s periodicity: the (standard) FS with and without interference errors, the TBUA heuristic with interference errors, and for the baseline RR are presented in Figure 6. One can observe oscillations that are correlated with the SINR variations (see Figure 5). Most of the time, the TBUA varies between the FSs and the RR scheduling. The curves for the FSs and TBUA solutions are most of the time above those of the RR scheduling. As stated above, the impact of interference errors is negligible, which is an important characteristic of the FS based solutions.

Fig. 6. Time variation of users' throughput for a scheduling period of $T=2s$, averaged over a sliding window of 15 iterations.

Figure 7 compares the MUT gain with respect to the baseline RR for the ten users, averaged along the full trajectory crossing the cell, for different FS scheduling periodicities. For 2s periodicity, the MUT gain for the TBUA and FS is of 15 and 30 percent respectively, and it grows monotonically to 80 and 117 percent respectively for 20s periodicity. One can observe a close to linear monotonic growth of MUT gain till 15s scheduling periodicity, with non-significant improvement brought about the long term time-space diversity beyond this periodicity.

VI. CONCLUSION

The allocation of FS involves the solution of a convex optimization problem. The solution complexity scales up with the number of users and the scheduling duration. In view of reducing the computational complexity of the FS we have proposed a two step heuristic in which, for

Fig. 7. FS MUT gain with and without interference errors, and TBUA wrt the RR scheduler as a function of the scheduling period (in sec).

each scheduling period, we select the two users with the highest α -fair utility and then we apply to them the FS. It has been shown that for $n = 2$, the FS allocation can be written in a closed form, making the heuristic even more computational efficient. Numerical simulations show that the throughput gain achieved by the FS heuristic remains significant, even for short scheduling periodicity.

REFERENCES

- [1] H. Zaaraoui *et al.*, "Forecast scheduling for mobile users," in *IEEE International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, 2016.
- [2] —, "Forecast scheduling and its extensions to account for random events," in *21st Innovations in Clouds, Internet and Networks (ICIN 2018) Conference*.
- [3] W. Hapsari *et al.*, "Minimization of drive tests solution in 3GPP," *IEEE Communications Magazine*, vol. 50, no. 6, pp. 28–36, 2012.
- [4] B. A. Fette, *Cognitive radio technology*. Academic Press, 2009.
- [5] Y. Zhao *et al.*, *Network support—The radio environment map*. Elsevier.
- [6] J. Ojaniemi *et al.*, "Optimal field measurement design for radio environment mapping," in *IEEE 47th Annual Conference on Information Sciences and Systems (CISS)*, 2013, pp. 1–6.
- [7] A. Konak, "A kriging approach to predicting coverage in wireless networks," *International Journal of Mobile Network Design and Innovation*, vol. 3, no. 2, pp. 65–71, 2009.
- [8] J. Riihijarvi *et al.*, "Characterization and modelling of spectrum for dynamic spectrum access with spatial statistics and random fields," in *IEEE 19th International Symposium on Personal, Indoor and Mobile Radio Communications*, 2008, pp. 1–6.
- [9] J. P. Chilès and P. Delfiner, *Geostatistics, Modeling Spatial Uncertainty*, 2nd ed. New-York: John Wiley & Sons, 2012.
- [10] N. Cressie *et al.*, "Fixed rank kriging for very large spatial data sets," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 2008.
- [11] H. Zaaoui *et al.*, "Analytical results for two users' forecast scheduling," November 2017. [Online]. Available: <https://hal.inria.fr/hal-01633361>
- [12] I. CVX Research, "CVX: Matlab software for disciplined convex programming, version 2.0," <http://cvxr.com/cvx>, Aug. 2012.
- [13] S. Boyd *et al.*, *Convex optimization*. Cambridge university press, 2004.
- [14] A. Tall *et al.*, "Virtual sectorization: design and self-optimization," in *5th International Workshop on Self-Organizing Networks (IWSON 2015)*, Glasgow, Scotland, May 2015.
- [15] V. Blondel *et al.*, *Recent advances in learning and control*. Springer, 2008.