End-to-end resource management and service quality control with mobility in LTE / LTE-Advanced networks

Hind Zaaraoui

UNIVERSITÉ D'AVIGNON
ET DES PAYS DE VAUCLUSE

THÈSE

DOCTEUR de l'**Université d'Avignon et des pays de Vaucluse**

pour obtenir le grade de
Spécialité: **Informatique et mathématiques appliquées**

27 mars 2017

présentée par
Hind ZAARAOUI

préparée au sein de Orange Labs et l'Université d'Avignon et des pays de Vaucluse dans le cadre de l'**École Doctorale 536: Informatique**

Gestion de ressources et contrôle de la qualité de service en mobilité dans les réseaux LTE/LTE Advanced

Thèse soutenue publiquement devant le jury composé de:

Tijani CHAHED	Telecom SudParis	Président
Samson LASAULCE	L2S, CentraleSupelec	Rapporteur
Urtzi AYESTA	IRIT	Rapporteur
Dr. Eitan ALTMAN	INRIA	Directeur
Dr. Zwi ALTMAN	Orange Labs	Co-directeur
Dr. Tania JIMENEZ	LIA	Co-directeur

To my parents, Saloua TAG and Ali ZAARAOUI

Acknowledgements

Abstract

The technology evolution of Radio Access Network (RAN) in the context of 5th Generation (5G) is not only guided by improving the network performance but also by the need to transform all the technologies into intelligent dynamic ones. The new 5G is a flexible technology that will be able to satisfy at the same time each user of any type of mobility (static or mobile) or service request (real and non-real time service) without modifying any models or algorithms in networks. All physical use cases will be able to be considered by the network intelligently and resource managed automatically.

The objective of this thesis is to analyse and enhance radio performance taking into account vehicular mobility by managing dynamically and intelligently the available resources.

To this end, we describe different users mobility models for discrete and continuous modeling. The discrete model using the well-known *car following* model is well adapted for simulations. The continuous one is useful to derive analytical key performance indicators (KPI). The novelty of the thesis is the analytical formulation of KPIs that take into account the physical mobility in the radio traffic which is not necessary stationary. As an example, the impact of a traffic light on performance indicators in a cell is investigated. It is shown that a periodical physical traffic congestion due to the traffic light deteriorate periodically the cell performance.

A first given solution is to improve resource allocation and control in the context of LTE-Advanced heterogeneous network. A small cell is deployed near the traffic light to relieve periodic congestion and QoS degradation. Three resource allocation and control schemes are investigated: a full frequency reuse, a static and a dynamic frequency splitting algorithm that are optimized with respect to a throughput based α-fair utility.

For sake of financial and energy costs decreasing, another solution is provided using new antenna array technologies in order to manage efficiently heterogeneous, fixed and mobile traffic. A heterogeneous antenna system with different large antenna array technologies is considered to offload static congestion areas and also the dynamical mobile congestion: Virtual Small Cell (VSC), virtual small cell with Self-Organizing Network (VSC-SON) and beamforming with multilevel global codebook that manages the heterogeneous antenna system at the Base Station (BS). The first two technologies improve the cell performance due to the capability to focus the signal at the traffic concentration. The novel beamforming solution with global codebook can further and significantly improve performance due to the capability to focus the signal along the road and to implicitly balance the traffic between the different

antennas. We compare all these technologies and their impact on the network performance.

The issue of user selection to allocate a portion (in time or in bandwidth) of the available resource is also analyzed. Moreover the context of resource management and network performance for 5G in high mobility is one of the future challenges. Thanks to the Minimization of Drive Testing (MDT) technology, networks can have Signal to Interference plus Noise Ratio (SINR) information with Geo-Localized Measurements (GLM). We introduce the concept of Forecast Scheduler for users in high mobility. It is assumed that a Radio Environment Map (REM) can provide interpolated SINR values along the user trajectories. Mobile users experience in their trajectories different mean SINR values. In mobile networks, schedulers exploit channel quality variation by giving the signal to the user experiencing best channel conditions while remaining fair. Nevertheless, we cannot record data rates of users with high mobility due to a very small time coherence. The Forecast Scheduling will exploit the SINR variation during users' trajectories. This scheduling is formulated as a convex optimization problem namely the maximization of an $\alpha-$fair utility function of the cumulated rates of the users along their trajectories. The optimal solution of this problem is given by using CVX algorithm. However this algorithm takes long time to solve the problem. An analytical solution and an algorithm are introduced in this thesis for the case of 2 users and an algorithm using this solution is given for the general case. The Forecast Scheduling model is also extended to take into account different type of random events such as arrival and departure of users and trajectories uncertainty.

Résumé

L'évolution technologique de RAN dans le contexte de la 5G n'est pas seulement guidée par l'amélioration de la qualité de service du réseau mais aussi par la nécessité de transformer toutes les technologies par des systèmes dynamiques intelligents. La nouvelle technologie de la 5G est plus que flexible et pourra satisfaire chaque utilisateur de façon équitable sans que son type (mobile ou statique) ou sa demande de service (service temps réel et non réel) n'affecte. Tous les cas d'utilisation seront intelligemment dimensionnés et gérés dans le réseau.

L'objectif principal de cette thèse est d'analyser et d'améliorer les performances radio en tenant compte de la mobilité des véhicules en gérant dynamiquement et intelligemment les ressources disponibles. À cette fin, nous présentons différents modèles de mobilité des utilisateurs dans le cas discret et continu. Le modèle discret utilisant le modèle bien connu du car following est bien adapté pour les simulations. La méthode continue est utile pour obtenir des indicateurs analytiques de performance clés (KPI). La nouveauté de cette partie de la thèse est la formulation analytique de KPIs qui tiennent compte de la mobilité physique dans le trafic radio. A titre d'exemple, l'impact d'un feu rouge sur les indicateurs de performance dans une cellule est étudié. Il est montré qu'une congestion physique périodique du trafic due au feu de circulation détériore périodiquement la performance de la cellule.

La première solution considérée consiste à améliorer l'allocation et le contrôle des ressources dans le contexte du réseau hétérogène LTE-Advanced. Une *small cell* est ensuite déployée à proximité du feu rouge pour diminuer la congestion périodique et la dégradation de la qualité de service. Trois systèmes d'allocation et de contrôle des ressources sont étudiés: une réutilisation de fréquence complète, un algorithme de division de fréquence statique et dynamique qui sont optimisés par rapport à une utilité α -fair basé sur les débits reçus. En outre, il est montré que le système de commande dynamique est particulièrement intéressant pour le trafic non stationnaire comme celui introduit par un feu de circulation périodique.

Par souci de réduction des coûts financiers et énergétiques, et de suivi d'utilisateurs mobiles, une autre solution est fournie en utilisant un nouveau réseau d'antennes afin de gérer efficacement le trafic hétérogène, fixe et mobile. On considère trois technologies différentes de réseau d'antennes permettant de décharger les zones de congestion ainsi que la congestion mobile dynamique: Virtual Small Cell (VSC), petite cellule virtuelle avec réseau auto-organisateur (VSC-SON) et rayons focalisant avec un multiniveau *global codebook* qui gère

le système d'antenne hétérogène à la station de base. Les deux premières technologies améliorent la performance de la cellule en raison de leur capacité de focaliser le signal à la concentration du trafic au niveau de la congestion physique. La nouvelle solution de faisceaux focalisants avec le global codebook améliore de manière significative les performances en raison de la capacité de focaliser le signal le long de la route et d'équilibrer implicitement le trafic entre les différentes antennes. Nous comparons toutes ces technologies et leur impact sur les performances du réseau.

La question de la sélection d'un utilisateur pour lui allouer une partie (en temps ou en bande passante) de la ressource disponible est étudiée. Le contexte de la gestion des ressources et de la performance du réseau dans la mobilité est un des défis futurs pour la 5G. Grâce à la technologie MDT, les réseaux peuvent avoir une idée, ou plus précisément une moyenne du SINR grâce à la GLM. Nous introduisons le concept de Forecast Scheduling pour les utilisateurs en mobilité à vitesse élevée. On suppose que la REM peut fournir des valeurs interpolées du SINR le long des trajectoires des utilisateurs. Les utilisateurs mobiles expérimentent dans leurs trajectoires différentes valeurs moyennes du SINR. Dans les réseaux mobiles, les ordonnanceurs exploitent la variation de la qualité du canal en donnant le signal à l'utilisateur qui connaît les meilleures conditions tout en restant équitable. Néanmoins, nous ne pouvons pas enregistrer le débit de manière instantanée des utilisateurs à mobilité élevée en raison d'un très faible temps de cohérence. Le Forecast Scheduling exploitera la variation de SINR sur les trajectoires futures des utilisateurs. Cette allocation est formulée comme un problème d'optimisation convexe, à savoir la maximisation d'une fonction d'utilité $\alpha-$ fair des débits des utilisateurs le long de leurs trajectoires. La solution de ce problème est donnée en utilisant l'algorithme CVX. Cependant, cet algorithme prend beaucoup de temps pour donner une solution. Une solution analytique est introduite dans cette thèse pour le cas des utilisateurs 2 et un algorithme utilisant cette solution est donné pour le cas général. Le modèle de Forecast Scheduling est également étendu pour tenir compte des différents types d'événements aléatoires tels que les arrivées et les départs d'utilisateurs ainsi que l'incertitude des trajectoires.

Contents

List of Abbreviations

3GPP: 3rd Generation Partnership Project

5G: 5th Generation

BS: Base Station

CQI: Channel Quality Indicator

eNB: enhanced eNode B

FDD: Frequency Division Duplex

FqS: Frequency Split

FRK: Fixed Rank Kriging

FS: Forecast Scheduling

FTT: File Transfer Time

GIS: Geographical Information Systems

GLM: Geo-Localized Measurements

GPS: Global Positioning System

GSM: Global System for Mobile Communications

HetNet: Heterogeneous Network

HSDPA: High Speed Downlink Packet Access

i.i.d: independent and identically distributed

KKT: Karush-Kuhn-Tucker

KPI: Key Performance Indicator

LoS: Line of sight

LSAS: Large Scale Antenna System

LTE: Long Term Evolution

MANET: Mobile Ad-Hoc Network

MB: Multilevel Beamforming

MDP: Markov Decision Process

MDT: Minimization of Drive Testing

MFqS: Mean Frequency Split

MIMO: Multiple Input Multiple Output

MMF: Max-Min Fair

MO: Macro cell Only

MUT: Mean User Throughput

OFDMA: Orthogonal Frequency-Division Multiple Access

p.d.f: probability density function

PF: Proportional Fair

PRB: Physical Resource Block

PS: Packet Scheduling

QoE: Quality of Experience

QoS: Quality of Service

RAN: Radio Access Networks

REM: Radio Environment Map

RFS: Restricted Forecast Scheduling

RNC: Radio Network Controller

RR: Round Robin

RS: Relay Station

RRC: Radio Resource Control

RRM: Radio Resource Management

RV: Random Variable

SFS: Seer Forecast Scheduling

SINR: Signal to Interference plus Noise Ratio

SON: Self-Organizing Network

TBUA: Two Best Users Algorithm

TDD: Time Division Duplex

TDMA: Time Division Multiple Access

TOP: Throughput Outage Probability

UFS: Updated Forecast Scheduling

UE: User Equipment

V2I: Vehicle-to-Infrastructure

V2V: Vehicle-to-Vehicle

VANET: Vehicular Ad-Hoc Network

ViS: Virtual Sectorization

VSC: Virtual Small Cell

List of mathematical notations

$\max_{x \in X} f(x)$: the maximum value of the function f in the set X

$argmax_{x \in X} f(x)$: the point x in the set X where f is maximal

$\mathbb{E}(Y)$: the expected value of the random variable Y

ds : the infinitesimal unit surface

$\mathbb{R}$: the set of real numbers

$\mathbb{N}$: the set of integers

$\frac{\delta f}{\delta x}$: the partial derivative of f with respect to x

$\Gamma(x)$: the Gamma function i.e. $\Gamma(x) = \int_0^{\inf} t^{x-1} e^{-t} dt$

Chapter 1

Introduction

1.1 Context

Smarter than the earlier generation (2G, 3G, 4G), the 5G of mobile networks goes beyond simply improving the network performance indicators (peak throughput, spectral efficiency, and latency). The main objectif of the 5G, in an unprecedented manner and particularly in the RAN, is to integrate *dynamicity, flexibility* and *modularity* in the new technologies.

- *Flexibility*: the new network system should have the capability to update its functionality for each use case by implementing different solutions (technologies and algorithms);
- *Dynamicity*: the new network will be able to adapt the radio features and virtual RAN architecture to any new context including mobility;
- *Modularity*: the new network may decompose formerly integrated network elements into more fine-grained functions.

5G networks have to be resource and cost efficient, easily implementable and operable. The flexibility and the dynamicity of the 5G networks should allow to integrate easily any novel technology of 5G radio access (beamforming technology, dynamical schedulings, Self-Organizing Network (SON)...) for any new use case. To this purpose, one of the most important subject of 5G is to adapt automatically the technology to each use case.

Providing seemless mobility in cellular networks is one of the challenges in cellular networks as much as improving the radio resource management and network performance for 5G in high mobility. The new architecture of

the 5G virtual RAN is expected to facilitate variety of options for both the deployment location of a virtualised RAN function and selections connectivity modes (multi-connectivity).

Besides the fact that mobility is a key topic in 5G RAN, the reason for the important regain of interest in mobility modeling in networks is twofold. The first is related to Heterogeneous Network (HetNet) in Long Term Evolution (LTE)-Advanced with low power nodes, such as small cells and femtocells. HetNets are seen today as a promising capacity solution for the permanent growth in traffic demand. A more accurate modeling of mobile traffic behavior will help to plan the network, and to design efficient control, resource allocation in high-mobility and self-optimization algorithms in physical congestion case for example. The second motivation is related to 5G networks and the corresponding scenarios which are currently under definition and study. In the Metis FP7 European project for example, the *Metis Horizontal Topics* have been defined for 5G, including the corresponding technologies for different scenarios [1]. Among these are the Ultra Dense Networks (UDN) and the Moving Networks (MN) for which fine modeling of mobility is clearly a challenge.

Mobility has been considered in the framework of Vehicular Ad-Hoc Network (VANET) which is a particular case of Mobile Ad-Hoc Network (MANET). The VANET technology considers vehicles as mobile nodes. Each vehicle can connect either to another vehicle (Vehicle-to-Vehicle, V2V) or directly to the infrastructure (Vehicle-to-Infrastructure V2I) ([2], [3],[4]).

This thesis considers vehicular users without any VANET technology and the communication between the infrastructure and the user is established through the mobile phone, the laptop or the connected tablet.

1.2 Objectives

The main goal of this thesis is to provide solutions for efficiently managing mobility/high-mobility in the network in the context of LTE advanced and 5G. We identify four objectives with relation to mobility in the network:

Objective 1: Create new dynamic key performance indicators

The first goal of this thesis is to model the Key Performance Indicator (KPI)s for network Quality of Service (QoS). To this end, vehicular mobility models and KPIs in static case are presented in order to introduce new dynamical QoS indicators that describe the impact of mobility on networks. A particular case of congestion in the context of mobility (traffic light) is defined to reveal this impact.

Objective 2: Analyze the impact of mobility in networks and resource management in case of traffic congestion

The second objective of the thesis is to provide solutions for managing the congestion problem in networks. One solution is to integrate small cells inside the macro cells. Small cells are then deployed next to the physical congestion area. The problem of the interference between the macro cell base station and the small cell one leads to implement solutions to manage the interference issue. Dynamical resource sharing methods between the base stations inside the cell are therefore used.

Objective 3: Provide resource management solutions for users in mobility in the presence of antenna system with beamforming capability

In order to avoid more costs on base stations' deployment and more energy consumption, one can capitalize on the offered central base stations and their antennas. M-MIMO antennas can be used to focus all the energy on some particular users. Beamforming technology and especially a low cost (in time and energy consumption) focusing beams are then integrated to the network system to serve mobile and static users.

Objective 4: Improve scheduler performance for high mobility by making use of GLM along the user trajectory

Users with high mobility do not benefit from channel gains due to Doppler effect and fading problems. In general, scheduling in older generations' technology (2G, 3G, 4G) only take advantage of short term time diversity of the

channel and do not address this issue. In this thesis, the envisioned solution is to provide a new scheduling model to respond to the high-mobility problem. We utilise GLM in order to benefit from long term time and spatial diversity of users along their trajectories. We assume that a REM can provide interpolated SINR values along the user trajectories.

1.3 Contributions

The main contributions of this thesis are to provide technological technics, models and algorithms to manage the resource between base stations and optimize resource allocation between users in case of mobility and high-mobility.

Previous analytical works involving mobility suppose that the users number is stationnary and compute the mean performance indicators in time and space. The novelty of this work investigates mobility with non-stationary pattern in the presence of different technologies and models that:

- Offload the loaded macro cells due to traffic congestion;
- Enhance quality of service indicators for mobile and fixed users;
- Follow the *mobile congestion* i.e. a group of mobile users;
- Improve the quality of service of mobile users by taking advantage on their *mobility*.

To this end, we have firstly created new analytical models of network performance that takes into account physical mobility which allow to control the performance and quality of service evaluation in the cell in different use cases: dynamic cell load, dynamic mean user throughput, dynamic file transfer time and outage probability in time (Chapter 3). To validate these models and to analyse the impact of the mobility on the network, a particular case of physical congestion is integrated in the system: the traffic light scenarios (Section 4.1).

In order to offload the macro cell due to traffic light congestion, a solution of small cell deployment next to the physical congestion is proposed in this work (Section 4.2). A dynamical resource sharing between the macro and small cells is integrated to eradicate the interferences issue and to consider the massive handover of mobile users between macro and small cells. Another way to offload the congestion area is to focus a virtual small cell in this area using the MIMO antennas technology. Even better, a low cost solution using M-MIMO antenna is considered to follow users leaving the congestion locations and enhance the network performance (Section 4.3). The solution consists on using a predefined multilevel codebook of beams. This technique is called **Multilevel Beamforming**.

Another issue is analyzed in this thesis: resource sharing between users in mobile and static mode. Previous scheduling models consider the fairness only for static or low velocity users. However users with high velocities were not taken into consideration and their high variation of fading cannot be *detected*.

We propose in this thesis a new way to schedule all type of users and particularly the users in high mobility. We call this new scheduler the **Forecast Scheduling** (Chapter 5). In an unprecedented manner of resource allocation, this sharing approach takes advantage on the past, present and future *received* data rate using GLM in order to exploit long term time and spatial diversity in the scheduling process of mobile users.

1.4 Publications

Hind Zaaraoui, and Zwi Altman. "Alleviating cellular network congestion caused by traffic lights." *IEEE Vehicular Technology Conference* (VTC Spring), 2015 IEEE 81st.

Hind Zaaraoui, Zwi Altman, and Eitan Altman. "Beam focusing antenna array technology for non-stationary mobility." *IEEE Wireless Communications and Networking Conference* (WCNC), 2016.

Hind Zaaraoui, Zwi Altman, Eitan Altman and Tania Jimenez. "Forecast scheduling for mobile users." *IEEE Personal, Indoor, and Mobile Radio Communications* (PIMRC), 2016 IEEE 27th Annual International Symposium on.

Hind Zaaraoui, Zwi Altman, Eitan Altman and Tania Jimenez. "Forecast scheduling and its extensions to account for random events". Submitted to Elsevier computer networks.

Chapter 2

Physical vehicular mobility model

"Resource Management and service quality control in context of ***MOBILITY*** *in LTE/LTE advanced networks."*

The main concern on studying mobility is to give the best resource allocation to users in mobility considering their radio condition. Mobile users can be pedestrian, vehicular, train, bus and/or plane travelers. All users using their mobile phones, laptops, connected tablets to download, communicate, etc. while moving from position A to position B are considered mobile users.

A pedestrian user is a user that has a high random mobility behavior. One cannot predict his trajectory neither his speed. On the other hand, a train, a bus or a plane traveler has a deterministic mobility behavior. Vehicular users have also a random mobility behaviour but are the most interested users to study as they can *learn* and *predict* their speeds and trajectories. In this thesis we choose the vehicular users as the mobile users in the studied network.

Mobility behavior and patterns have been studied in different scientific communities, dealing with vehicular traffic, or crowd behavior in mass gathering events [5]. In the case of vehicular traffic, the dynamics can be captured by microscopic (discrete) models which describe the velocity and acceleration of the individual cars, taking into account the interactions between neighboring cars. Continuous models describing the behavior of cars' motion can also be considered, and are of interest for deriving analytical solutions for performance indicators.

In this chapter we present a state of the art of different models of vehicular mobility. A simulation model for mobile users used through the manuscript (except for the last chapter of the Forecast Scheduling were users have a fixed

speed) is described in the last section. A computation of vehicles' distribution in case of continuous and discrete cases is proposed.

2.1 Discrete case

In the network, each user is considered as a single person who needs a resource to download a requested data. The discrete mobility models describe the evolution of each user speeds and trajectories. We propose two different type of discrete mobility: the random one and the deterministic one. For the random mobility models (not very realistic, especially for the vehicular users), four models from literature are listed below:

- Random waypoint mobility model;
- Random walk mobility model;
- Gauss-Markov model;
- Boundless simulation area mobility model.

Deterministic discrete models are also introduced and are more realistic for vehicular mobility:

- Car following model;
- Cellular Automaton model.

Our simulation model for vehicular mobility is inspired from the Car Following model presented in Section 2.3.

2.1.1 Random models

Random waypoint mobility model

This model was proposed by Johnson and Maltz [6] and it is often used as a benchmark to evaluate mobile ad hoc network (MANET) routing protocols, because of its simplicity.

In this model, between each motion step, the user marks a pause with a determited duration (t_{pause}). During the motion step, the mobile node selects randomly a velocity v between a lower and a higher speeds (in general the lower speed is null), and travels to a random selected position (x, y) between the offered ones. Note that if $t_{pause} = 0$ this leads to a continuous motion.

The random waypoint mobility model is generally used in simulations for analyzing the impact of mobility on the KPIs in different fields.

Random walk mobility model

Also called the Brownian motion mobility model or Brownian walk, this model gives to mobile users the possibility to move at each fixed time step dt with a random direction and speed from predefined interval $[0, 2\pi]$ and $[v_{min}, v_{max}]$. The random walk was initially proposed to study the random motion of particles in physics. It has some similarities with the random waypoint mobility model especially when $t_{pause} = 0$. However, these two models differ in the speed variation duration. For the random waypoint the variation are in space change and in the random walk in time step change i.e. in the random walk model users speed change during each duration time step dt which is not the case in the random waypoint model where the speed will change as soon as the user arrives at a defined position.

In case of spatial boundaries, the mobile node bounces off the border with the angle determined by the arrival direction impact.

Random Gauss-Markov model

This model was first introduced by Liang and Haas [7]. Mobile user's speed in this model is correlated over time and uses the Gauss Markov stochastic process. Denote by $v(t)$ and $d(t)$ the speed and the direction at time step t. At each time t, the mobility parameters equations are defined as follows:

$$v(t+1) = \alpha v(t) + (1-\alpha)V + \sqrt{(1-\alpha^2)}W^v,$$
$$d(t+1) = \alpha d(t) + (1-\alpha)D + \sqrt{(1-\alpha^2)}W^d,$$

where W^v and W^d are uncorrelated random Gaussian process with mean 0 and variance σ^2; V and D are the mean values of respectively speed and direction; and α is the tuning parameter used to vary the randomness of the model with $1 \geq \alpha \geq 0$. Speed and direction become totally random if $\alpha = 0$ and linear if $\alpha = 1$. Level of randomness is obtained by varying α.

Boundless simulation area mobility model

In this model, unlike the previous ones, there exists a relation between previous state (last speed and last position) and present one (similarly to the random Gauss-Markov model). In every time step dt, we update the mobility parameters speed and direction as following:

$$\begin{aligned} v(t+dt) &= \min(v_{max}, \max(v(t)+dv, 0)), \\ \theta(t+dt) &= \theta(t) + d\theta, \\ x(t+dt) &= x(t) + v(t)cos(\theta(t)), \\ y(t+dt) &= y(t) + v(t)sin(\theta(t)), \end{aligned}$$

where dv and $d\theta$ are selected randomly with uniform distribution from speed and angle predetermined intervals.

When a node reaches the boundary space of the simulation, it does not bounce back but it continues his travel and reappears on another random side of the simulation.

2.1.2 Deterministic models

Car following model

The principle of this model is that every vehicular user n behaves *sensibly* identically as the vehicle ahead $n-1$ (see Fig.2.1):

$$response = sensitivity \times stimulus.$$

The car following model, the simpliest one known as Chandler model(1958) [8], can be written as following:

$$\frac{dv_n(t+T)}{dt} = a'(v_{n-1}(t) - v_n(t)),$$

where a' is the sensitivity coefficient.

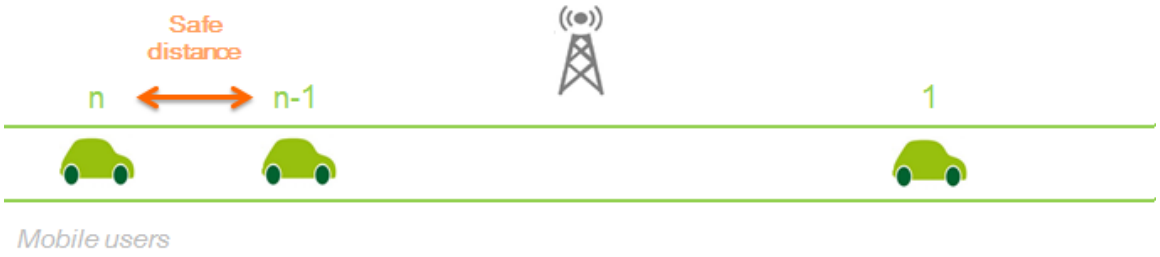


Figure 2.1: Distribution of vehicles in the case of the car following model

Another non-linear car following model developed by Gazis (1961) [9] is known as the General Motors non-linear model and is defined as following:

$$\frac{dv_n(t+T)}{dt} = a' \frac{v_n(t)^m}{(x_{n-1}(t) - x_n(t))^l}(v_{n-1}(t) - v_n(t)),$$

where l is the sensitivity coefficient of the distance, m is the sensitivity coefficient of the speed and $x_n(t)$ is the position of the n^{th} vehicle.

In the car following model, one can add a *safe distance* between cars (see Fig.2.1). The safe distance car following model describes the relation between two following cars. The most basic model is the Pipes one wich says: "A good rule for following another vehicle at a safe distance is to allow yourself at least the length of a car between you and the vehicle ahead for every ten miles an hour ($16.1km/h$) of speed..."[10]. By this definition, the safe distance that vehicle n must have to the vehicle $n-1$ ahead is described as following:

$$D_n(v) = L_n(1 + \frac{v}{16.1}) \tag{2.1}$$

where L_n is the length of the n^{th} vehicle.

Other models integrate the decision time and the braking time needed to apply the brakes, risk factor (Leutzbach 1988 [11], Jepsen 1998 [12], etc.).

Cellular Automaton model

Cellular automaton model is one of the discrete vehicular traffic models the most used in simulations. This model assumes that vehicles move from one cell to another (Nagel and Schreckenberg) i.e. the street is divided into cells of lengths which depend on the car length (see Fig.2.2). Each cell can be

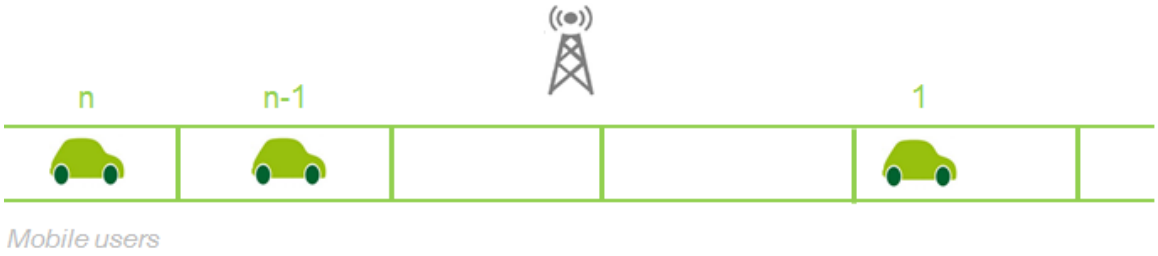


Figure 2.2: Distribution of vehicles in case of the Cellular Automaton model

occupied by at most one car. The basic rule is that each mobile user moves C cells at each time depending on its speed. Each unity of speed adds up to the distance of one unit cell if there is no car ahead. However if there is a car ahead with a distance less than v speed then the car moves to the cell just before the car ahead:

$$x_n(t+1) = x_n(t) + \min(v_n(t), x_{n-1}(t) - x_n(t) - 1) \tag{2.2}$$

where v_n is updated as follows:

$$v_n(t+1) = \max(0, \min(v_{max}, x_{n-1}(t) - x_n(t) - 1, v_n(t) + 1) - \omega_n) \tag{2.3}$$

ω_n is a Boolean random variable that can be added to integrate the decelaration case where $\omega_n = 1$.

2.1.3 Mean vehicles number and distributions

We compute in this section the mean vehicles number in an interval position $[x, x + dx]$. The speed model considered can be one of the models described previously.

Denote by $\lambda(t)$ the arrival intensity of cars in a studied network cell. Suppose that at the border of the cell the number of arriving vehicles is a Poisson process. Define $Z(x,t)$ as the random variable of the number of vehicles between x and $x + dx$, N_t the number of vehicles in the entire cell that have arrived between 0 and t. $r(0, T_n, t)$ is the location at time t of the n-th vehicle entered in the cell at position 0 (cell border) at time T_n:

$$Z(x,t) = \sum_{n=1}^{N_t} 1_{r(0,T_n,t)\in[x,x+dx]}. \tag{2.4}$$

Denote by D_1 the expectation value of Z. The j-th time of vehicular user arriving has $\Gamma(j, \frac{1}{\lambda})$ as distribution (called also Erlang law). Recall that $r(0, T_n, t) = r(T_n, t)$ the position of the vehicule at time t, arrived at time T_n into the system at position 0.

$$D_1(x,t) = \mathbb{E}(\sum_{i=1}^{\infty} 1_{r(T_i,t)\in[x,x+dx]} 1_{T_i \leq t})$$

$$D_1(x,t) = \sum_{i=1}^{\infty} \mathbb{E}(1_{r(T_i,t)\in[x,x+dx]} 1_{T_i \leq t})$$

$$D_1(x,t) = \sum_{i=1}^{\infty} \int_0^t 1_{r(u,t)\in[x,x+dx]} e^{-u\lambda} \frac{u^{i-1}}{(i-1)!\frac{1}{\lambda}^i} du$$

$$D_1(x,t) = \int_0^t 1_{r(u,t)\in[x,x+dx]} e^{-u\lambda} \sum_{i=1}^{\infty} \frac{u^{i-1}}{(i-1)!\frac{1}{\lambda}^i} du$$

thus

$$D_1(x,t) = \lambda \int_0^t 1_{r(u,t)\in[x,x+dx]} du \tag{2.5}$$

It is more difficult to find an explicit expression for $D_1(x,t)$ if λ depends on time. This case is generated when there is a deceleration or acceleration of vehicles (for example vehicles stopped by a traffic light). To overcome this problem, one can assume that λ is a staircase function in time and proceed the calculation of D_1 at each duration time where λ is constant.

The discrete model tends to the continuous one when the distance between following cars tends to zero. In the next section, we present continuous models based on fluid dynamics.

2.2 Continuous case

The main aim of studying the continuous models is to integrate them in the analytical performance indicators of networks. Two models are given from the literature:

- Gas-Kinect traffic flow model;
- LWR model.

2.2.1 Gas-Kinect traffic flow model

The gas-kinect flow model gives a description of the heterogenity of traffic flow. This model treats vehicles as particles of gas. The Boltzmann equation for traffic flow has been initially used by Prigogine and Herman in 1971 [13] to compute the velocity distribution function $f(x, v, t)$:

$$\frac{\delta f(x,v,t)}{\delta t} + v\frac{\delta f(x,v,t)}{\delta x} = -\frac{f(x,v,t) - \rho(x,t)F_{des}(v)}{\tau_{rel}} + (\frac{\delta f(x,v,t)}{\delta t})_{int},$$

where

$$\begin{aligned}(\frac{\delta f(x,v,t)}{\delta t})_{int} = &\int_{w>v}(1-p^*(\rho))|w-v|f(x,v,t)f(x,w,t)dw \\ &-\int_{w<v}(1-p^*(\rho))|v-w|f(x,v,t)f(x,w,t)dw,\end{aligned}$$

where τ_{rel} is the relaxation time and $F_{des}(v)$ is the desired velocity distribution.

2.2.2 LWR model

The Lighthill-Whithman-Richards (LWR) model is an approach that supposes that the expected velocity v depends on the density function of mobile users p i.e.:

$$v(x,t) = v^e(p(x,t)).$$

The partial differential equation that describes the relation between mobile users distribution and their speed is as follows:

$$\frac{\delta p}{\delta t} + \frac{\delta p v^e(p)}{\delta x} = 0.$$

Solutions of this equation are not unique and can be determined by using the characteristic curves along which a initial information is diffused (Leutzbach 1988 [11]).

2.2.3 Distribution of vehicles

Consider next the computation of the mean number of vehicles $p(x, t)$ at each space and time unit interval. Suppose a flow of vehicles arrives at position and time $(0, 0)$, namely the entering time of the first vehicle in the cell at its border. $v(x, t)$ evolves with a continuous traffic model (LWR or Gas-Kinect

traffic flow model, or any other model). The position or the distance r of that flow will evolve as follows:

$$\frac{dr}{dt} = v(r(t), t). \tag{2.6}$$

$\lambda(u)$ cars arrive with an initial speed V_0 at time $u < t$ and their speed varies with a continuous model. Denote by dx the interval unit where at least (for instance) the SINR is *"constant"*. If at time t these λ cars are between x and $x + dx$, we add λ to the number of vehicles existing in this space interval. Recall that $r(0, u, t)$ is the position of the flow at time t that arrived at time u from location 0. Then mean number of vehicles is described as following:

$$p(x, t) = \int_0^t \lambda(u) 1_{r(0,u,t)\in[x,x+dx]} du \tag{2.7}$$

2.3 Mobility Model for simulations

A deterministic behavior of mobile users is considered in our simulator. As each vehicular user responds to the behaviour of the vehicle in front of him, we consider in this case the car following model to describe users' speed. For sake of simplicity, we propose a new car following model where only overtaking is excluded.

The inter-arrival of vehicles in a cell is supposed to be an exponential random variable with parameter λ that can depend on time. The first vehicle enters the cell area with an initial speed V_0. When it approaches to any source of physical congestion with a distance close to a safety distance it begins its deceleration linearly. The following vehicle begins its path with the same initial speed V_0. As it approaches the front vehicle with a safety distance, the vehicle starts decelerating too. Denote by n the n-th car arriving to the cell, and by T the reaction/reflection time. We implemente our car following algorithm which preserves their relative order (overtaking is excluded), i.e. $\forall t$:

$$x_i(t) + v_i dt < x_{i-1}(t) + v_{i-1} dt \tag{2.8}$$

where i is the i^{th} car arriving in the cell, v_i is its speed and x_i is its position. There exists a parameters $a' < 1$ (we take for all this thesis $a' = 0.8$) such that (2.8) can be reformulated as follows:

$$v_i = a'(v_{i-1} + \frac{x_{i-1} - x_i}{dt}) \tag{2.9}$$

In the simulator, we use hereafter a car following algorithm for n vehicles on a road in a cell which preserves the relative order of the cars. The algorithm used in the simulator is given by Table 2.1:

Table 2.1: Car following algorithm

$$\begin{cases} x_1 = 0, v_1 = 1 \text{ at } t = 0\text{; } 1^{st} \text{ user initial conditions ;} \\ x_1, x_2, ..., x_n \text{ are given at time } t\text{;} \\ v_1 \text{ dynamic evolution at time } t+1\text{;} \\ \text{for } i = 2 \text{ to } n \text{ do:(evaluation of } v_i \text{ one by one)} \\ v_i = a'(v_{i-1} + \frac{x_{i-1} - x_i}{dt}); [dt = 1] \\ \text{end} \\ \text{update all } x_i \leftarrow x_i + v_i dt \end{cases}$$

We rewrite the speed of the i^{th} vehicle $v_i(t) = v_i(x_i, t)$,

Chapter 3

Performance and QoS indicators for static and dynamic networks

*"Resource Management and **SERVICE QUALITY** control in context of mobility in LTE/LTE advanced networks."*

To improve the QoS and the mobile user experience in networks, one should define first of all the appropriate KPIs and then optimize them for different use cases. As users in mobility are dynamic, the KPIs must depend on time. Previous works analyze the impact of mobility considering queuing theory and Markov chain. Other works involving mobility suppose that the users number is stationnary and compute the mean performance indicators in time and space. This chapter provide a new way to study the performance indicators on dynamic network based on the static KPIs.

We first introduce different indicators in static case (users are not mobile):

- Data rate, peak rate and mean data rate;
- SINR outage probability;
- Throughput outage probability;
- Cell load;
- File transfer time;
- Cell capacity.

New dynamical indicators, depending on users' speed, trajectories and intensity in time, are deduced from the static ones:

- Mean user throughput;
- Dynamic traffic and cell load;
- Dynamic SINR outage probability;
- File trasfer time depending on users' mobility behaviour.

Random mobility models and particularly random waypoint (section 2.1.1) have been introduced into VANET to investigate the impact of the mobility on network performances ([14], [15]). A more realistic mobility model, the STRAW (Street Random Waypoint) model has been considered in [16], and its impact on performance (e.g. packet delivery ratio) has been analyzed through simulations.

In [17], Whittle model is used to evaluate the mobility effect on cellular network. This model assumes that each user remains fixed until he finishes his data transmission. The Wireless Elastic Traffic (WET) model considers users that transmit during their travel. Mobility model in this case is a stationary Markov process where users can move around the cell discontinuously in time and space [18]. As the analytical solution for the WET model is difficult to derive, Whittle approximation and queuing theory are needed to derive the mean value of performance indicators (e.g. of file transfer time or mean user throughput).

As users speed may vary in time, the number of users in each unit area is not stationary. For example, a red light or an accident can lead to a congestion in some areas. In the non-stationary case, instantaneous KPIs such as throughput outage probability or mean instantaneous users' throughput are of interest. In this work we derive analytical formulas for instantaneous performance indicators taking into account physical mobility of users. To this end we combine microscopic models for vehicular mobility that capture the dynamics and interaction between cars with radio modeling. Analytical modeling taking into account physical mobility allows one to accurately evaluate performance and quality of service in the cell.

3.1 Performance indicators and resource management for static users

3.1.1 Data rate, peak data rate and mean user throughput

In all this section we consider a cell with downlink transmissions that shares a resource *fairly* between users depending on their data rate. Denote by $a_i(t)$ the fraction of time or bandwidth resource allocated to user i with $\sum_{i=1}^{n} a_i(t) = 1$ with n being the number of users in the cell. The data rate of a user i in position x_i is defined as follows:

$$R'_i(x_i) = a_i(t)R(x_i), \tag{3.1}$$

where $R(x_i)$ is the peak data rate which means the data rate received for user i if he were alone in the cell. Note that if $a_i(t) = 1$ then $R'_i(x_i) = R(x_i)$ in accordance with this definition. If a Round Robin (RR) scheduling is used as resource sharing method in the cell then $a_i(t) = \frac{1}{n(t)}$, where $n(t)$ is the number of users at time t in the cell.

Using the Shannon formula, the peak data rate $R(x_i)$ is defined as a function of SINR in position x_i:

$$R(x_i) = Wlog_2(1 + SINR(x_i)). \tag{3.2}$$

where W is the total available bandwidth. The peak data rate can also be defined depending on the user's position from the antenna (angles θ_i and ϕ_i see Section 4.3.2) and his distance to the base station i.e. $R(x_i) = R(r_i, \theta_i, \phi_i)$ where r_i is the distance of user i to the base station.

The Signal to Interference plus Noise Ratio (SINR) incorporates the fast fading F which should be taken into account when users are mobile. A rayleigh fading is assumed, namely an exponential random variable with parameter 1 $F \sim exp(1)$. Shadowing S (log normal random variable) and path loss $r^{-\alpha}$ are also considered in the model plus thermal noise N. Denote by $j = 1...J$ the BSs numbers and by F_i and S_i the fast fading and shadowing of station i respectively. r is the distance of a user to the serving BS and r_j the distance to the $j - th$ interferer. The aggregate interference power at the receiver is defined by:

$$I = \sum_{j=1}^{J} F_j S_j r_j^{-\alpha} \tag{3.3}$$

The SINR is then defined as follows:

$$SINR = \frac{F_0 S_0 r^{-\alpha}}{N + I}. \tag{3.4}$$

where F_0 and S_0 are respectively the fading and the shadowing of the serving BS.

Hence, the mean user throughput (MUT) is defined as follows:

$$MUT = \frac{1}{n} \sum_{i=1}^{n} R'_i(x_i). \tag{3.5}$$

3.1.2 SINR Outage probability

We make the following assumptions to simplify the next calculations of the outage probabilities:

- Noise can be neglected compared to the total interference in a (dense) urban environment
- Fenton-Wilkinson model is used to approach the sum of log-normal random variables with a log-normal RV [19].
- The fast fading can increase or decrease the received power. We suppose that the increase of interfering power due to fast fading coming from some base stations are compensated by the decrease of interfering powers coming from others. Therefore we assume that $F_j \approx \mathbb{E}(F_j) = 1$.

SINR outage occurs at distance r to the BS when SIR or SINR is less than a threshold δ. The SINR outage probability is then $\epsilon = P(SIR < \delta)$.

$$\epsilon = 1 - P(F_0 S_0 r^{-\alpha} > \delta \sum_{j=1}^{J} F_j S_j r_j^{-\alpha}).$$

With the above-mentioned assumptions we have:

$$P(F_0 S_0 r^{-\alpha} > \delta \sum_{j=1}^{J} F_j S_j r_j^{-\alpha}) \approx$$

$$P(F_0 S_0 r^{-\alpha} > \delta \sum_{j=1}^{J} S_j r_j^{-\alpha}).$$

For the calculation of this probability we use the definition of conditional probability, the assumption that the fading is $F \sim exp(1)$ and The Fenton-Wilkinson approximation which include normal distribution with a mean m and standard deviation σ:

$$\epsilon = \int_0^\infty Q(\frac{10\log_{10}(x/\delta) - m}{\sigma})e^{-x}dx \tag{3.6}$$

where $Q(u) = \frac{1}{2}erfc(\frac{u}{\sqrt{2}})$ the mean m and the standard deviation σ are defined as follows:

$$m = \frac{10}{ln(10)}ln(\frac{\sum_{j=1}^{n} r_j^{-\alpha}}{r^{-\alpha}K(r,\alpha)})$$

$$\sigma^2 = 2(1 - (\frac{10}{ln(10)})^2 ln(K(r,\alpha)))$$

where K is defined as following:

$$K(r,\alpha) = e^{a^2/2}((e^{a^2/2} - 1)\frac{\sum_{j=1}^{n} r_j^{-2\alpha}}{(\sum_{j=1}^{n} r_j^{-\alpha})^2} + 1)^{-1/2}$$

note $a = \frac{ln10}{10}$. The integral (3.6) can be written as an expectation of an exponential random variable function with $Y \sim exp(1)$:

$$\epsilon = \mathbb{E}(Q(\frac{10\log_{10}(Y/\delta) - m}{\sigma})). \tag{3.7}$$

3.1.3 Throughput outage probability

The throughput outage probability of a cell is the mean of the outage probability per *surface* unit. This KPI is the probability that the throughput Thp experienced by the users located in the given surface unit does not exceed some threshold δ'. The expression for this probability is $\epsilon' = P(Thp < \delta')$. We make the assumption that the throughput is a function of the SINR according to the Shannon formula and a RR scheduler. Let n be the mean number of users at time t and denote by $R'(x,t)$ the throughput:

$$R'(x,t) = W\frac{log_2(1 + SINR)}{n} \tag{3.8}$$

W being the bandwidth. Thus we have the throughput outage probability written as a probability that the SINR does not exceed a certain threshold which depends on the number of users:

$$\begin{aligned}\epsilon' &= P(Thp < \delta') \\ &= P(SINR < 2^{\frac{\delta' n}{W}} - 1).\end{aligned} \tag{3.9}$$

Using the formula of the SINR outage probability, we can deduce the analytical expression of throughput outage probability ϵ':

$$\epsilon' = \mathbb{E}(Q(\frac{10\log_{10}\left(\frac{X}{2^{\frac{\delta' N}{W}}-1}\right) - m}{\sigma})). \tag{3.10}$$

For simulation of outage probabilities, the Monte Carlo method is used. Assume that $(y_1, y_2, ..., y_k)$ are independent variables $\sim exp(1)$, then the convergence is almost sure:

$$\lim_{k\to\infty} \frac{1}{k}\sum_{i=1}^{k} Q(\frac{10\log_{10}(y_i/\delta) - m}{\sigma}) = \epsilon'. \tag{3.11}$$

3.1.4 Cell load

We suppose that the traffic demand is uniformly distributed in the cell. In case of fixed users, data flows arrive as Poisson process with intensity λds for a surface unit ds. Each user arrives with a data volume σ random independent and identically distributed (i.i.d.), then the traffic density per surface unit is defined as follows:

$$\rho = \lambda\mathbb{E}(\sigma). \tag{3.12}$$

The traffic intensity generated in area of distance of $[r, r + dr]$ to BS (see Fig.3.1) is defined as follows:

$$d\rho(r) = \rho ds, \tag{3.13}$$

ds is defined by $2\pi r dr$ (see Fig.3.1).

The load generated in this surface unit corresponds to the users demand on the network resource and is defined by the ratio $\frac{d\rho(r)}{R(r)}$ [20]. The total cell load is the sum of this load all around the surface cell (with *radius* G):

$$\overline{\rho} = \int_0^G \frac{d\rho(r)}{R(r)}. \tag{3.14}$$

3.1.5 File transfer time

The file transfer time of a user i is the duration which this user is served by the network. This service time depends then on user's data volume σ requested and

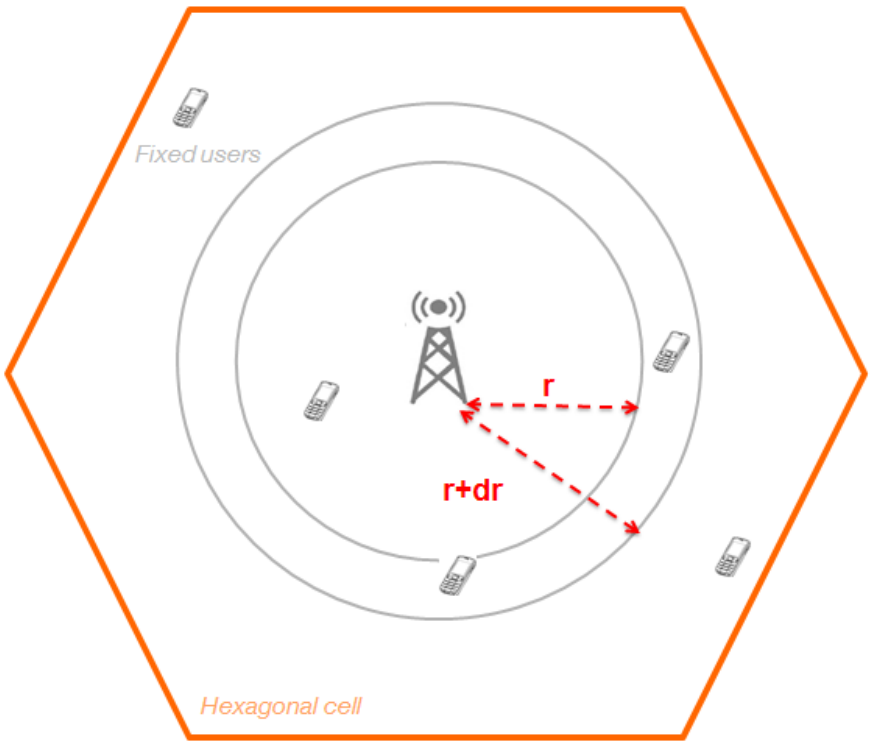


Figure 3.1: The hexagonal cell model with the infinitesimal unit distance dr definition

his peak data rate $R(r_i)$ when the user is alone. The File Transfer Time (FTT) of user i is defined as follows:

$$FTT_i = \frac{\sigma}{R(r_i)}. \tag{3.15}$$

In the continuous case, we define the typical service time distribution in the cell by:

$$dFTT = \frac{\sigma}{R(r)} \frac{2rdr}{G^2} \tag{3.16}$$

3.1.6 Cell capacity

This parameter describes the maximum traffic intensity for which the cell is not congested i.e. $\overline{\rho} < 1$.

As the available resource is supposed to be fairly shared between users, users arrive with a Poisson processus with intensity $\lambda\pi G^2$ in the cell. The cell

load is then written as following:

$$\overline{\rho} = \lambda \pi G^2 \times \int_0^G \frac{\mathbb{E}(\sigma) 2rdr}{R(r)G^2}. \tag{3.17}$$

Using the formula of the cell load (3.14), we have the following equality:

$$\overline{\rho} = \lambda \pi G^2 \times \int_0^G \frac{\mathbb{E}(\sigma) 2rdr}{R(r)G^2} = \int_0^G \frac{d\rho(r)}{R(r)}. \tag{3.18}$$

We deduce from this equality the cell capacity formula:

$$C(G) = (\int_0^G \frac{2rdr}{R(r)G^2})^{-1}. \tag{3.19}$$

3.2 Dynamic performance and QoS indicators for mobile users

In this section we introduce main dynamic QoS performance indicators: mean user throughput, dynamic traffic and cell load, and file transfer time for mobile users.

3.2.1 Mean user throughput MUT

Recall that $R'(x,t)$ is the throughput (Thp) that is achieved in position x at time t and by $MUT(t)$ the instantaneous mean user throughput at time t:

$$MUT(t) = \frac{1}{P(t)} \int_S R'(x,t)p(x,t)dx, \tag{3.20}$$

where $P(t) = \int_S p(x,t)dx$ is defined by the eq.2.7.

Or in discrete case:

$$MUT'(t) = \frac{1}{n} \sum_{i=1}^{n} R'(x_i(t),t). \tag{3.21}$$

3.2.2 Dynamic traffic and cell load

In this section, we use the notation introduced in section 3.1.4, namely, the offered traffic is denoted by ρ, the cell load by $\overline{\rho}$, the arrival intensity of data flows per surface unit dS by λ, and the flow data sizes arrived in this surface by σ which are independent and identically distributed.

The load density is $d\rho/R(x)$ where $R(x)$ is the peak data rate at position x and $d\rho = \lambda \mathbb{E}(\sigma)dS$.

We introduce a time varying offered traffic ρ due to mobility. Recall that the load can be defined as the proportion of time when users are in the cell. In the elastic traffic, we assume that all the BS resource is used even if there is only one user in the cell.

In section 2.2.3 we have already found an expression for the distribution densities $p(x,t)$ for all vehicles (continuous models) at each time. The physical arrival intensity $\lambda(x,t)$ is related to the density distribution $p(x,t)$ by the one dimension conservation equation:

$$\frac{\partial p(x,t)}{\partial t} + \frac{\partial \lambda(x,t)}{\partial x} = 0. \tag{3.22}$$

Remark: In wireless networks, λ is often used to denote arrival intensity of active (transmitting) users. λ is used here for arrival intensity of cars, which may or may not have active users : σ can be null for some users.

When a vehicle begins its transmission at (a, b) where a is the position and b - the time, it has a $\sigma(a, b)$ data volume to transmit along the road. We suppose in this work that the data volume for all users is constant. During its trajectory, the vehicle experiences different SINR at different locations. Denote a time unit by dt (1 sec. for the simulations). A vehicle that arrives with speed $v(x, t)$ traveled $dx = v(x, t)dt$. dt is supposed very small so that during the vehicle moving the SINR can be considered as constant at the interval $[x, x + dx]$. At each time unit dt the user with an initial data volume $\sigma(a, b)$ transmits a small data volume that depends on the time delay dt and its position x.

Denote by $\sigma_l(x, t)$ the data volume left from the beginning of the transmission (a,b) to its present location x at time t, and by - $A(x)$ the space interval $[x, x + dx]$. If the user begins its data transmission from position a at time b then the size of data volume that is transferred during dt at interval position $[x, x + dx]$is:

$$d\sigma(x, t) = 1_{a+\int_b^t v(r(a,b,s),s)ds \in A(x)} \min(R'(x, t)dt, \sigma_l(x, t)) \tag{3.23}$$

where $R'(x, t)$ is the data rate received at the position x at time t.

The indicator function in (3.23) allows to filter users that have arrived to the interval $[x, x + dx]$ at time t. The amount of data arriving at time t in the position x can be written as the instantaneous traffic density $d\rho(x, t)$ defined as: $d\rho(x, t) = \lambda(x, t)d\sigma(x, t)$. In the continuous case the instantaneous cell load is as follows:

$$\overline{\rho} = \int_S \frac{d\rho(x, t)}{R(x)} \tag{3.24}$$

where $R(x)$ is the peak data rate or in other words, the data rate received if there is only one user.

In the discrete case, each user (in the simulator) has an equivalent data volume to transmit at each iteration (1 *iteration* = 1 *second*). $d\rho(x_i, t) = d\sigma(x_i, t) = d\sigma_i(t)$. Hence the instantaneous cell load for discrete case is:

$$\overline{\rho} = \sum_{i=1}^{N_t} \frac{d\rho(x_i, t)}{R(x_i)}. \tag{3.25}$$

Note that $\overline{\rho}$ is either equal to 0 if there are no users, or 1 if all users are actually transmitting i.e. $min(R'(x_i, t), \sigma_l(x_i, t)) = R'(x_i, t)$. $\overline{\rho}$ is between 0

and 1 if there is at least one user i who is ending the data transfer during interval dt i.e. $min(R'(x_i,t),\sigma_l(x_i,t)) = \sigma_l(x_i,t)$. If we suppose that σ is infinite then $\overline{\rho}$ will be always equal to 1 since no user is leaving the network.

3.2.3 SINR outage probability

We define the dynamic outage probability of a cell at time t as a mean of the outage probabilities at each position weighted with the traffic density:

$$P_c(t,\delta) = \frac{1}{\overline{\rho}(t)} \int_S P(SIR(r) < \delta) d\overline{\rho}(t), \tag{3.26}$$

where $P(SIR(r) < \delta)$ is the ϵ defined by the formula (3.6). For the discrete case (for the simulations), we define the outage probability of the cell as the mean users outage probability:

$$P_c^*(t,\delta) = \frac{1}{n(t)} \sum_{i=1}^{n} P(SIR(x_i(t)) < \delta) \tag{3.27}$$

where $n(t)$ is the number of users and $x_i(t)$ is the position of the user i.

3.2.4 File transfer time

FTT in mobility (applicable also for fixed users) represents how much time a mobile user needs to transfer all his data as long as he is in the cell. The file transfer time for user i is given by :

$$FTT_i = min(\int_0^\infty ((\frac{dx_i(t)}{v_i(x_i(t))} 1_{v_i \neq 0} + 1_{v_i=0}) 1_{x_i<L}) dt, \int_0^\infty \frac{d\sigma_i(t)}{R'(x_i(t))}) \tag{3.28}$$

where $x_i(t)$ is the position of the user i at time t whithin the driving path of length L and that verifies the dynamic equation $\frac{dx_i}{dt} = v(x_i)$. $1_{v_i=0}$ in the eq.(3.28) means that we compute also the time where the mobile user is stopped (for example by a traffic light or by any physical congestion).

Recall that the mobile users start their data transfer from the left cell edge while entering in the macro cell area. The file transfer time here computes how much time a mobile user stays in this cell until he finishes file transmission (second part of the *min* in eq.(3.28)) or finishes to cross the cell (first part of

the *min* in eq.(3.28)). If the time needed to finish the data transfer in this cell exceeds the duration to cross the cell then the file transfer time is the period of crossing the cell.

The FTT_i for the fixed user case:

$$FTT_i = \int_0^\infty \frac{d\sigma_i(t)}{R'(x_i, t)}. \tag{3.29}$$

Note that the FTT of mobile and fixed users have the same expression if we assume that for the fixed users $v_i(x_i(t)) = 0$ for all t. Therefore $\int_0^\infty 1_{v_i=0} 1_{x_i<L} dt = \infty$ and then:

$$FTT_i = min(\infty, \int \frac{d\sigma_i}{R'(x_i(t))})$$

which is equivalent to the formula (3.29).

Chapter 4

Network congestion management

"Resource Management and service quality ***CONTROL*** *in context of mobility in LTE/LTE advanced networks."*

While moving, vehicular users experience different areas with different signal power and may be confronted to different congestions. During working days, a vehicular users is very likely to endure peak hours from 8 a.m. to 10 a.m. and from 5 p.m. to 7 p.m. The main reason of these significant congestions is not only the arrival intensity of vehicles, but also the traffic lights that obstruct the vehicles in a small area during peak hours.

The long waiting in vehicles pushes the conductors and the travelers to use their mobile phone and to connect to internet, to call, to send messages etc. In general, a physical congestion in the roads leads to congestion in the network. Operators should therefore provide solutions to offload such congested cells.

In this chapter, we introduce a particular congestion which is the congestion due to traffic light and the impact of their position (either near or far from the base station in the cell) on network performance. The results are analyzed using numerical simulations. Three solutions are proposed to offload the congestion in the macro cell:

1. Small cell deployment next to the congestion area;
2. Virtual small cell covering the congestion area;
3. Multilevel beamforming to track the mobile users.

A comparison between the two last propositions are given in the numerical result in section 4.3.6.

4.1 Impact of vehicular mobility with traffic lights on wireless network performance

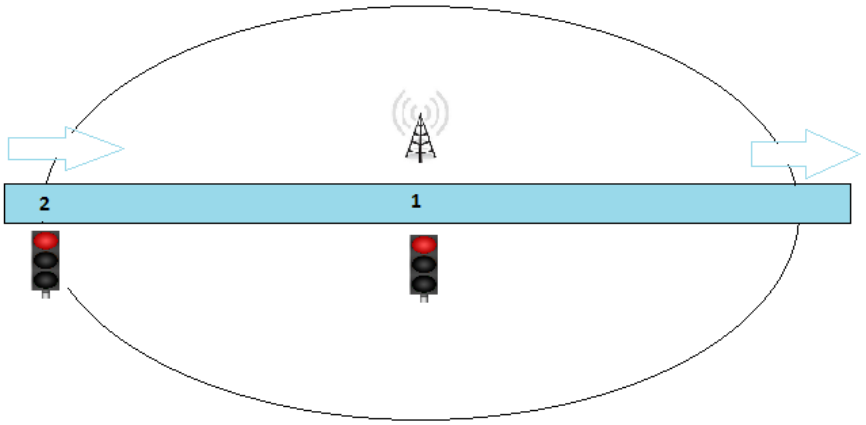


Figure 4.1: Cell with traffic lights scenarios

The particular case of a unique cell with a base station located at the center and a linear road crossing it is considered. We consider two scenarios: Scenario 1 has a traffic light located near the base station and scenario 2 has a traffic light at the cell edge. The cell is supposed initially empty. First users are coming from the left border of the cell with an infinite data volume and leave from the cell at the right edge (Fig.4.1). Vehicles are moving according to the car following model (Algorithm 2.1) defined previously. The simulations are performed using a Matlab simulator, with parameters given by Table 4.1. *One iteration (it.) in the simulator of this section is equivalent to 1 second.* Three performance indicators are considered: SINR and throughput outage probabilities and mean user throughput.

Traffic light near the Base Station

In real life, a traffic light takes in general a few seconds for the red light and vary depending on the location of the traffic light and also depending on the countries. In our simulations we consider a traffic light operating with a periodicity of 400 iterations, which is a long time but allows to enhance the physical congestion area. The impact of the periodicity can be seen in

Table 4.1: Network and Traffic characteristics

Network parameters	
Number of macro BSs	1
Number of interfering BSs	6
Macro Cell layout	hexagonal trisector
Intersite distance	$500m$
Bandwidth	$20MHz$
Channel characteristics	
Thermal noise	$-174dBm/Hz$
Macro Path loss (d in km)	128.1 + 37.6 $\log_{10}(d)$ dB
Small cell Path loss (d in km)	140.7 + 36.7 $\log_{10}(d)$ dB [21]
Mobility traffic characteristics	
λ_1 mobile users	0.5users/s
λ_2 fixed users	6users/s
Position traffic light case 1	next to the BS
Position traffic light case 2	left border of the cell
traffic light periodicity	200it. red/ 200it. green
Maximal user speed	$1m/it.$
File size for mobile users $sigma_1$	80 Mbits
File size for fixed users $sigma_2$	10 Mbits
SINR threshold δ	0.01
Throughput threshold δ'	0.1 Mbits

all the Figures below. The SINR outage probability (Fig.4.2) decreases when the light turns red as the distance between traffic lights and macro cell BS is small and the vehicles gradually accumulate near the base station (see roads 2 and 3 in Fig.4.5). The SINR outage probability increases as soon as the light turns green because the vehicles begin to move away from the BS/light (roads 1 and 4 in Fig.4.5). At the iteration 760 for example, there is a peak followed by a sharp decrease of the SINR outage probability. This behavior is due to the departure from the right cell edge of the accumulated vehicules that were previously blocked by the red light (see the car accumulation on the right border of the road 1 in Fig.4.5 which will leave the cell in the next iterations).

The growing number of users decreases the mean user throughput with time (see Fig.4.3). It is recalled that the cell is initially empty. The impact of the periodicity of traffic light is apparent in the three KPIs (presented in Figs.4.2-4.4). The outage behaviour is shown in Fig.4.4. It is noted that the mean outage may be acceptable while from user perspective, periodic degradation of

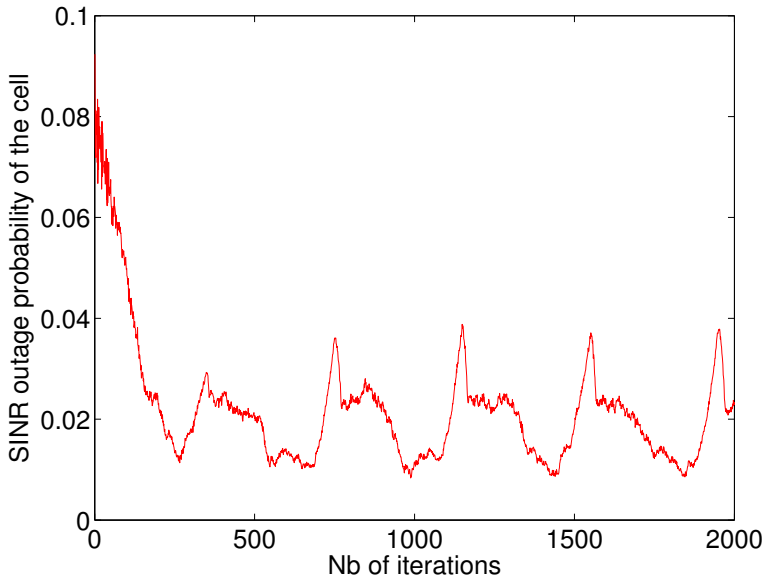


Figure 4.2: SINR outage probability of the cell. Traffic light located near the BS

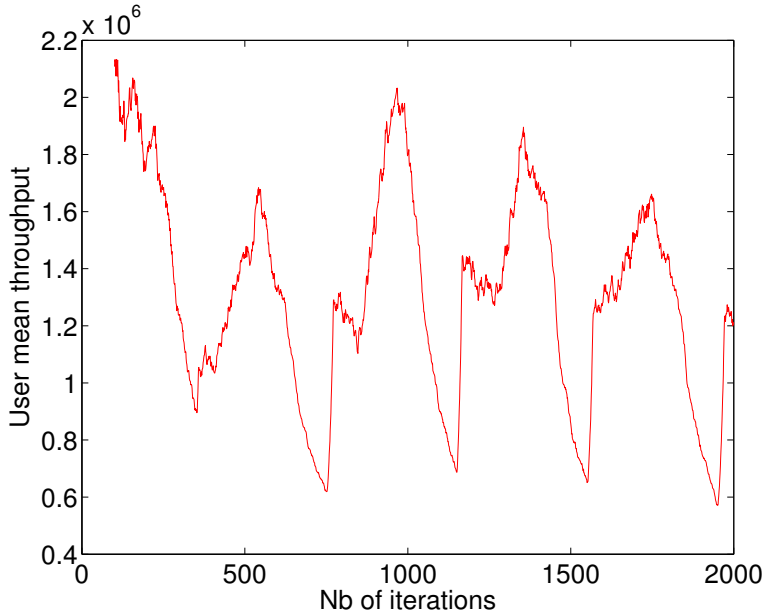


Figure 4.3: Mean user throughput of the cell. Traffic light located near the BS

outage may not.

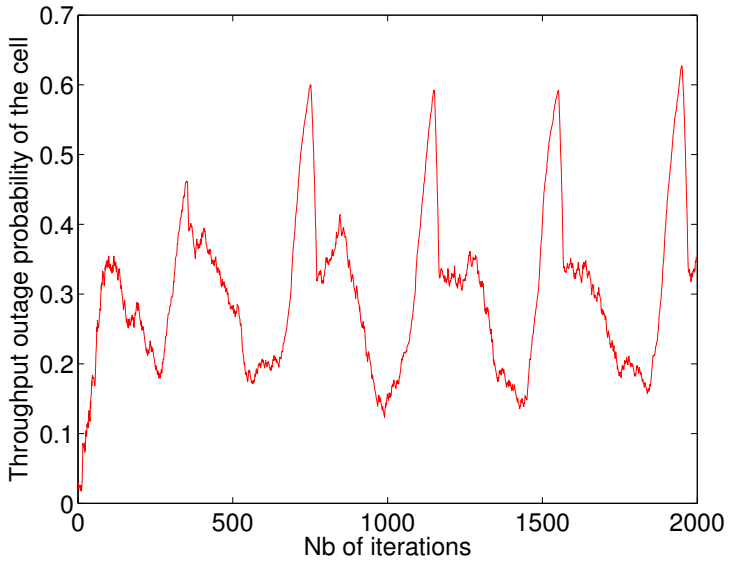


Figure 4.4: Throughput outage probability of the cell. Traffic light located near the BS

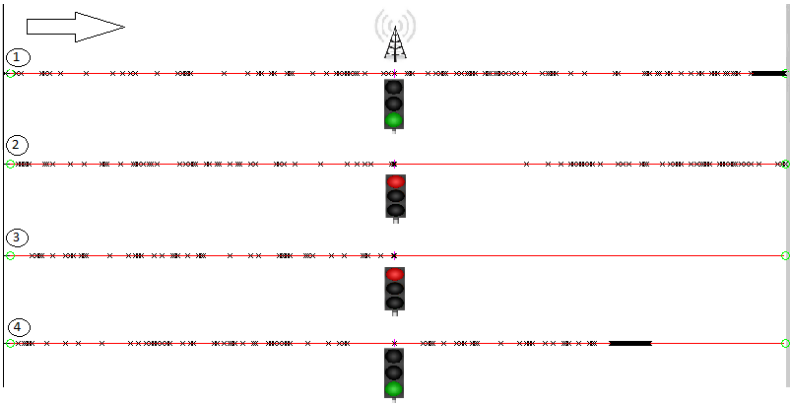


Figure 4.5: Vehicles distribution for iterations 760, 850, 950 and 1100

Traffic light near the cell border

In the second scenario, the traffic light is located near the cell edge. In this scenario the users during congestion are experiencing the worst service quality. At the first iterations, the light is red and vehicles must stop for the first 200 iterations. Fig.4.6 shows the SINR outage probability. It is fixed to the maximum value until the light turns green (200 iterations later). The probability then decreases due to the grouped vehicles approaching the BS. Then

the light turns red and an accumulation of a second group of vehicles near the cell edge increases the SIR outage probability. The small decrease (or discontinuity) seen during the global outage increase is due to the departure of the the grouped vehicles from the cell. The mean user throughput shown in Fig.4.7 starts by decreasing because of the growth of the number of users. At each traffic light period, peaks can be observed that are due to departures of groups of vehicles from the cell. The accumulation of vehicles is the result of cars stopping near the red light (see the snapshot of road 1 at iteration 900 in Fig.4.9). The corresponding results for the throughput outage probability are shown in Fig.4.8.

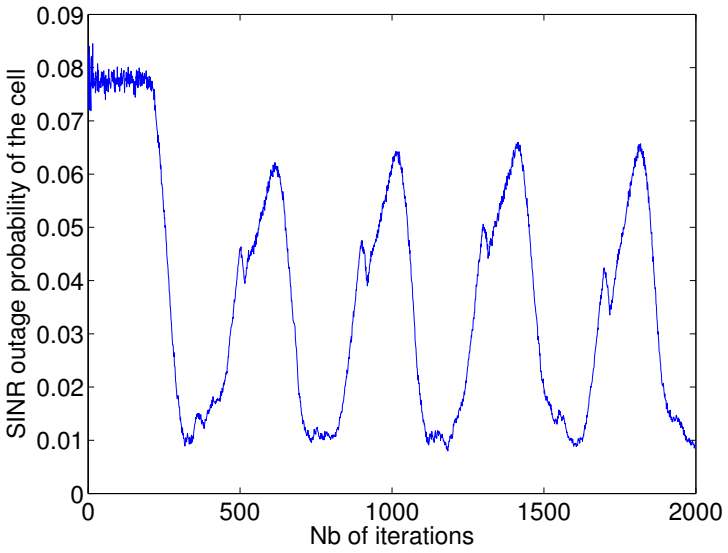


Figure 4.6: SINR outage probability of the cell. Traffic lights far from BS

Comparison between the two scenarios

When comparing the two scenarios we can see that the first difference is that the amplitudes of the SINR outage probability in scenario 1 is smaller (0.01 to 0.04) than in scenario 2 (0.01 to 0.065). This is due to the difference of distance of traffic lights to the BS. The second difference is the lag over time of the maximum and minimum peaks of the outage probability and user mean throughput. In fact, the maximum is reached for probabilities in scenario 1 for which vehicles move away (from the BS) in the case of green light. On the contrary, in the scenario 2 the maximum is attained during the red light with

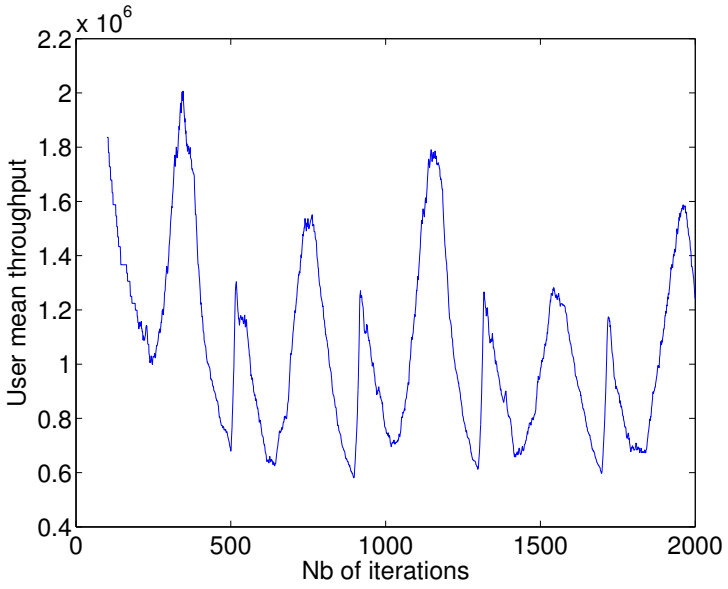


Figure 4.7: Mean user throughput of the cell. Traffic lights far from BS

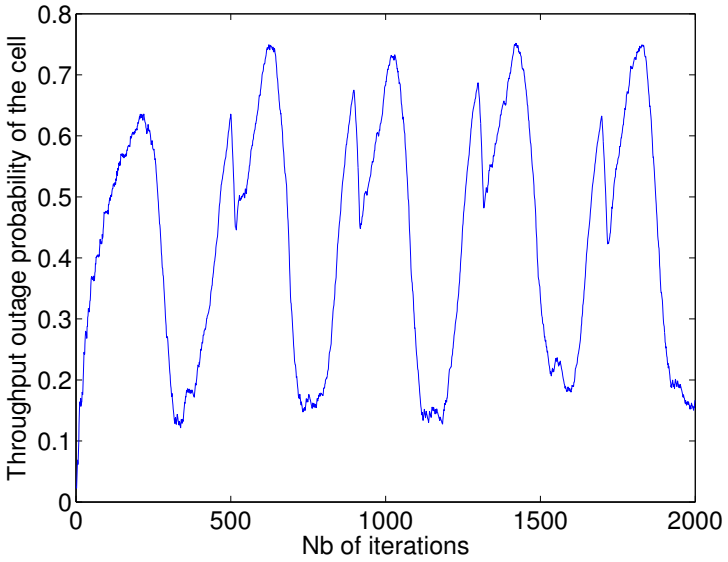


Figure 4.8: Throughput outage probability of the cell. Traffic light far from BS

users accumulating at the cell border. The mean user throughput reaches its maximum value at the red light for scenario 1 whereas it is at the minimum for scenario 2 for the same reasons. Notice also that the impact of the number

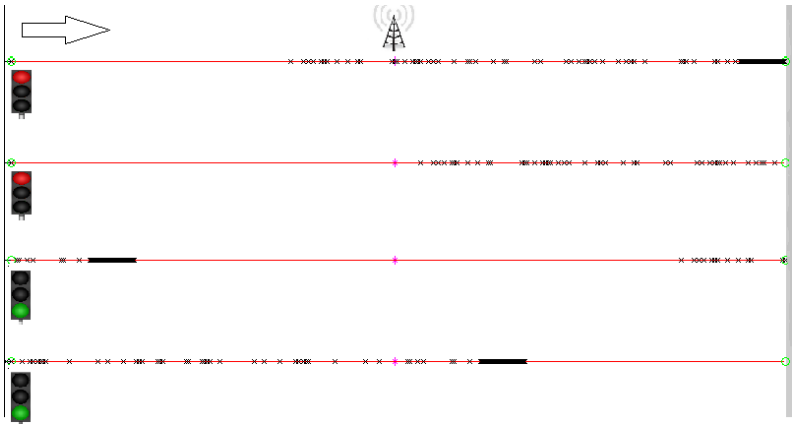


Figure 4.9: Vehicles distribution for iterations 900, 950, 1050 and 1200

of users is higher than the impact of the distance to the base station for the throughput outage probability or Mean User Throughput (MUT) (Figs 4.2 and 4.4 or Figs 4.6 and 4.8) during the first iterations.

Note that if we have studied only the mean in time of QoS performances, the instantaneous degradation and improvement of the KPIs would not be noticed. Integrate physical mobility in the instantaneous KPIs' models are important to describe the dynamic QoS of the network and particularly to analyze if dynamic resource sharing technologies are introduced to alleviate the network congestion.

Summary

Traffic light periodicity significantly impact QoS indicators such as SINR and throughput outage probabilities or mean user throughput that follow its periodicity. The physical congestion created in the cell edge deteriorate the users' KPI not only for the users in the congestion area but for all the users in the macro cell. It is shown that cell outage can periodically occurs due to cumulation of cars, even for medium level of traffic intensity.

To offload the macro cell of the cumulated users in a particular surface, namely at the cell edge next to the traffic light, the macro cell may need to be dimensioned and planned taking into account periodic QoS degradation due to mobility behaviour of users. This degradation may be hidden when considering average KPIs. The time variation of the KPIs due to user mobility provide rich information that can be used in control and (self-) optimization algorithms using small-cells, Massive MIMO antennas for virtual small cells and beamforming.

4.2 Resource sharing between small cell and macro cell

In this section, only scenario 2 with the traffic light near the cell edge is considered. We introduce a solution to offload the macro cell from the congestion due to periodic traffic light. The first solution comprizes a small cell that is deployed near the traffic light to relieve periodic congestion and QoS degradation.

By deploying a new small cell in the macro cell, the service quality of all the users in the macro cell is improved. However interferences are increased for all the cells in the network and especially between the macro cell and the small cell whithin its coverage. A way to mitigate the interference is to split the frequency bandwidth resource between the two cells.

The purpose of this section is to analyze different solutions to manage cell congestion and QoS degradation due to traffic light located far away from the macro-cell. Three resource allocation and control schemes are investigated: a full frequency reuse, a static and a dynamic [22] frequency splitting algorithm that are optimized with respect to a throughput based $\alpha-$fair utility [23]. The car following model introduced in section 2.3 is used to model the cars' speed and their interactions. Through numerical simulations, it is shown that the frequency splitting algorithms outperform the full frequency reuse scheme in terms of user throughput and file transfer time. Furthermore, it is shown that dynamic control scheme is of particular interest for non stationary traffic as the one introduced by a periodic traffic light.

When the network is densified by small cells, mobility patterns of mobile users such as group mobility can significantly impact the cell performance, congestion and the perceived QoS. The impact of traffic lights on macro-cell network performance has been studied in section 4.1.

A special focus is given to the control algorithm under non-stationary, periodic variations of the traffic introduced by the traffic light. While for stationary traffic distribution, fixed control parameters can provide optimal performance, it is shown here that for non-stationary case, dynamic control parameters considerably outperform the static one. This result shed light on the evaluation and choice between centralized off-line and on the importance of distributed on-line control and self-optimization solutions.

4.2.1 Resource allocation

The periodic accumulation of cars near the traffic light results in periodic congestion and outage of the macro cell. In order to relieve periodic congestion in the cell due to the traffic light located near the cell edge, we deploy a small cell near it. Three cases of bandwidth sharing between the macro and the small cell are analyzed: the full bandwidth reuse (Reuse), the optimal bandwidth sharing which takes into account the users' configuration at each iteration (FqS), and mean optimal bandwidth sharing (MFqS) which is based on optimizing the constant split factor for the average utility value. We consider next the three resource allocation schemes for releaving congestion and enhancing the cell capacity.

Full frequency reuse (Reuse)

Full frequency reuse consists of allocating the entire bandwidth to both the macro and the small cell. While benefiting from additional resources, these scheme suffers from mutual interference between the macro- and the small cell, and calls for interference mitigation techniques.

Dynamic frequency bandwidth splitting (FqS)

The frequency bandwidth is shared between macro-cell and small-cells in their coverage area. A split factor δ is optimized at each iteration as a function of the configuration of users. The main objectives of this method are to remove the interferences between macro cell and small cell and to split the frequency resources in a manner to optimize an α-fair utility of the users throughputs. Denote the surface of the small cell by S_s and the macro cell surface by S_m. The α-fair utility U_α of users throughputs that optimizes the bandwidth split factor δ in case of $\alpha \neq 1$ [22] is given by:

$$U_\alpha(\delta) = \sum_{x_i \in S_s} \frac{(\delta R'(x_i,t))^{1-\alpha}}{1-\alpha} + \sum_{x_i \in S_m} \frac{((1-\delta) R'(x_i,t))^{1-\alpha}}{1-\alpha}. \tag{4.1}$$

In the case where $\alpha \to 1$ (proportional fair), the utility is written as:

$$U_\alpha(\delta) = \sum_{x_i \in S_s} log(\delta R'(x_i,t)) + \sum_{x_i \in S_m} log((1-\delta) R'(x_i,t)). \tag{4.2}$$

The throughput $R'(x,t)$ is defined by Shannon formula which is a SINR function. We assume that there is one serving macro-cell and six interfering

macro-cells. The SINR for the macro cell m user and small cell s user at position x is defined as following:

$$SINR_m(x) = \frac{P_m h_m(x)}{N_0 + \sum_{k=1,k\neq m}^{7} P_k h_k(x)}. \tag{4.3}$$

$$SINR_s(x) = \frac{P_s h_s(x)}{N_0}. \tag{4.4}$$

In the case of the proportional fair (i.e. $\alpha = 1$) one has directly a closed form expression for the frequency split parameter δ. The optimal solution is therefore the solution of the equation $\frac{\partial U_\alpha(\delta)}{\partial \delta} = 0$:

$$\frac{\partial U_\alpha(\delta)}{\partial \delta} = \frac{N_s}{\delta} - \frac{N_m}{1-\delta} = 0 \tag{4.5}$$

where N_m and N_s are respectively the number of users in the macro cell and in the small cell. The solution of the above equation is then :

$$\delta = \frac{N_s}{N_s + N_m}. \tag{4.6}$$

Note that δ varies dynamically depending on N_s and N_m that change in time i.e. $\delta = \delta(t)$.

Mean optimal frequency bandwidth split (MFqS)

We have seen in the previous section that the frequency split factor must be optimized at each time iteration. Using the mean value of δ, denoted by $\overline{\delta}$, in time still outperforms the full reuse solution. Nevertheless, this fixed $\overline{\delta}$ does not fully optimize the mean utility over time. From simulation results, the $\overline{\delta}$ is 0.5 whereas the best optimal fixed δ^* is close to 0.6. We want to find a fixed δ^* that optimizes the mean α-utility in time for $\alpha = 1$. The mean α-utility in time is as follows:

$$\frac{1}{T}\int_0^T U_1(\delta^*) = \frac{1}{T}\int_0^T \sum_{x_i \in S_s} log(\delta^* R'(x_i,t)) + \sum_{x_i \in S_m} log((1-\delta^*)R'(x_i,t))dt. \tag{4.7}$$

By supposing δ^* fixed in time, we can take it out of the integral. This function is then derived with respect to δ^* and is equalized to 0 to find the optimal fixed parameter δ^*:

$$\delta^* = \frac{1}{1 + \frac{\frac{1}{T}\int_0^T N_m(t)dt}{\frac{1}{T}\int_0^T N_s(t)dt}}. \tag{4.8}$$

If the users number were stationary (if there are only fixed users or there is no traffic light), using the Ergodic theorem [24]:

$$lim_{T\to\infty}\frac{1}{T}\int_0^T N_{m/s}(t)dt = \mathbb{E}(N_{m/s}) \tag{4.9}$$

where $N_{m/s}$ the number of users in macro cell m or in the small cell s. This solution is different from the mean δ in time $\frac{1}{T}\int_0^T \frac{N_s}{N_s+N_m}dt$. From simulations, the mean number of users in the small cell is 3 and at the macro cell is 2.1. The optimal fixed δ^* is 0.588 which is the same result found by simulation.

This method is of interest in the periodic cases. The particular case of periodic congestion due to traffic light is the perfect example to implement this technic. One can compute two mean optimal frequency bandwidth split parameters: a δ_1^* when the traffic light is green and a δ_2^* when the light is red. However only a single δ^* is implemented in this work.

4.2.2 Simulation and results

The results for FTT, loads and mean user throughput are filtered with a sliding window of 20 iterations (seconds). One can see that the cell load with only the macro station (MO) is very high and almost equals to one (Fig.4.10). This shows that there is always at least one transmitting user and others ending the data transference. Mobile users are accumulated in a area where the SINR is low due to the traffic light located in bad channel conditions. Mean user throughput at each iteration time does not exceed $300Kbits/s$ in the presence of interfering base stations.

Comparing the three scenarios of bandwidth sharing and the reference macro-alone scenario, the most efficient one is the dynamic frequency bandwidth split solution. Figures 4.10-4.13 show how the load in each case evolves in time. Recall that these results are filtered with a sliding window of 20 iterations. In the full bandwidth reuse case (Fig.4.11), the load of the macro cell dropped to 0.35 thanks to the small cell that off-loads users accumulated near the traffic light. However the small cell load is almost equal to 1 except when users start leaving the small cell during green light. In order to share loads between cells, we have introduced the frequency bandwidth split solution. This

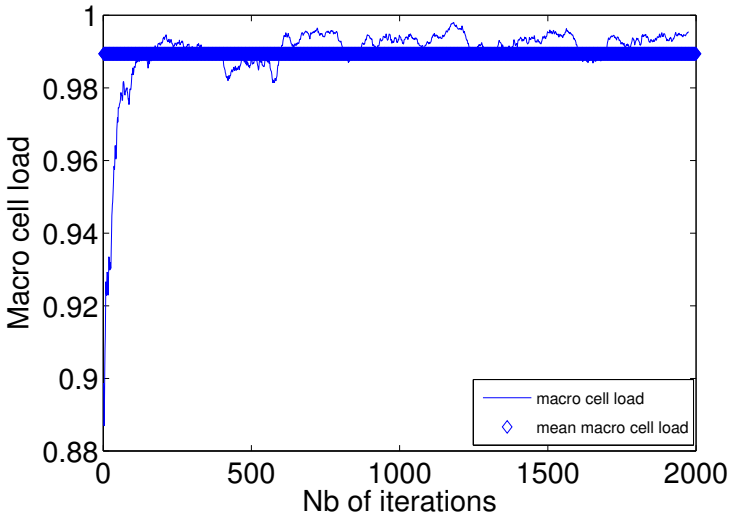


Figure 4.10: Macro cell load in the presence of the traffic light at the cell edge.

method allows to dynamicaly reduce high load values by optimally sharing frequency resources (Fig.4.12). Mean frequency split is also a good technic for balancing load and optimizing other performance indicators. Nevertheless it increases the macro cell load (Fig.4.13). If fixed and mobile users had the same data volume to transmit, one might have a higher average load gap between the dynamic δ-split and δ^*-split scenarios.

As we have a closed form of δ-split in case of $\alpha = 1$, the computing complexity time is pratically the same for the dynamic and for the mean optimal frequency bandwidth split.

Mean user throughput in time and mean file transfer time clearly show the impact of each bandwidth sharing scenario (see Fig.4.14 and Fig.4.15). File transfer time of the MO case is 6 times higher than that of the Reuse case, and the Reuse is 6 times higher than that of the FqS case.

From Fig.4.14 and Fig.4.15 one can see that the FqS solution provides a significant gain with respect to the MFqS, namely of 45% when taking MFqS as a reference, since it takes advantage of the periodic, non-stationary traffic characteristics.

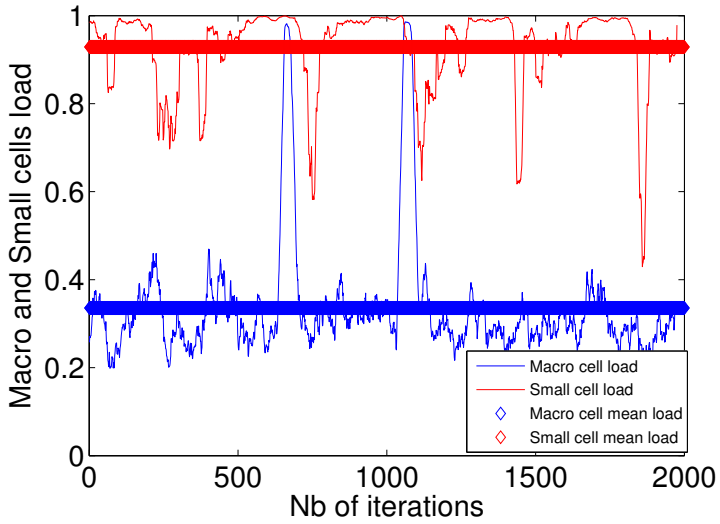


Figure 4.11: Macro and small cell loads in the presence of a traffic light at the cell edge. Full reuse bandwidth case.

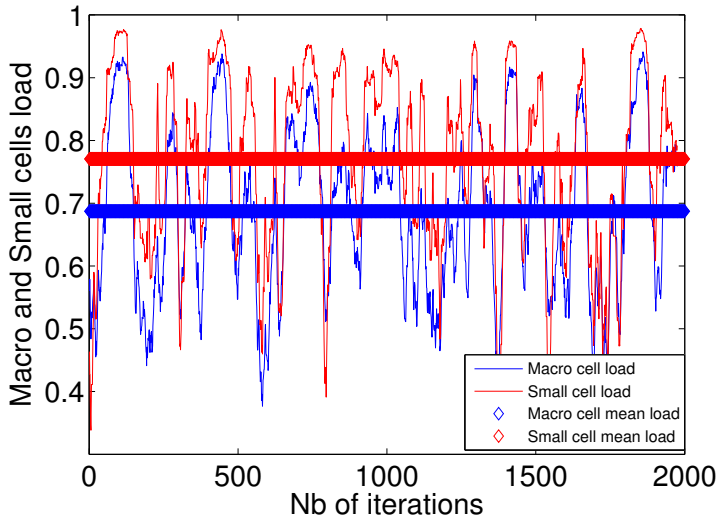


Figure 4.12: Macro and small cell loads in the presence of a traffic light at the cell edge. Frequency bandwidth split case.

Summary

We have considered the impact of vehicular mobility in the presence of traffic light. The traffic dynamics and the interactions between cars are mod-

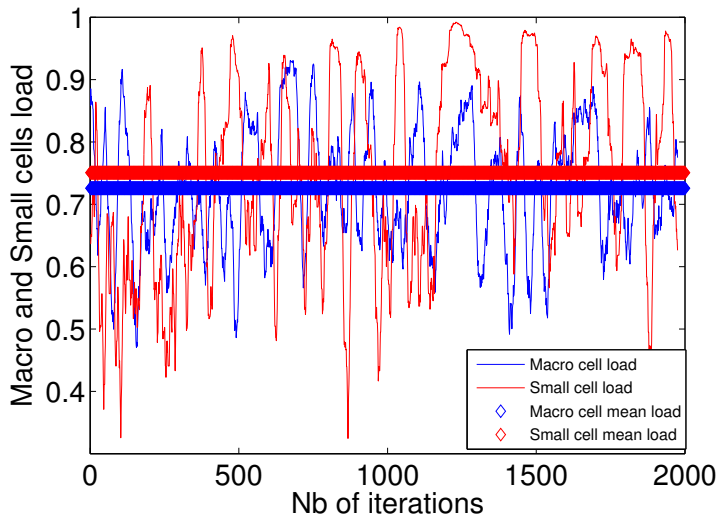


Figure 4.13: Macro and small cell loads in the presence of a traffic light at the cell edge. Mean frequency bandwidth split case.

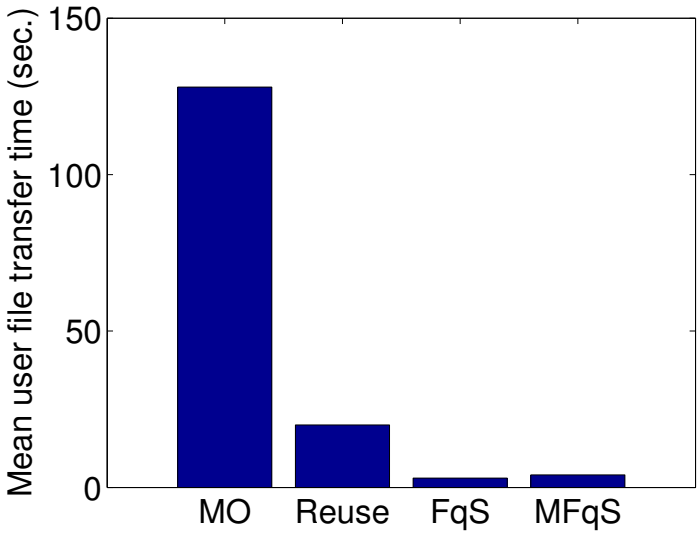


Figure 4.14: Mean user file transfer time in the cell for each bandwidth sharing scenario

eled using our particular car following model. The periodicity of the traffic light can introduce periodic degradation of the cell performance and perceived QoS. To alleviate congestion and periodic outage phenomena, a small cell has

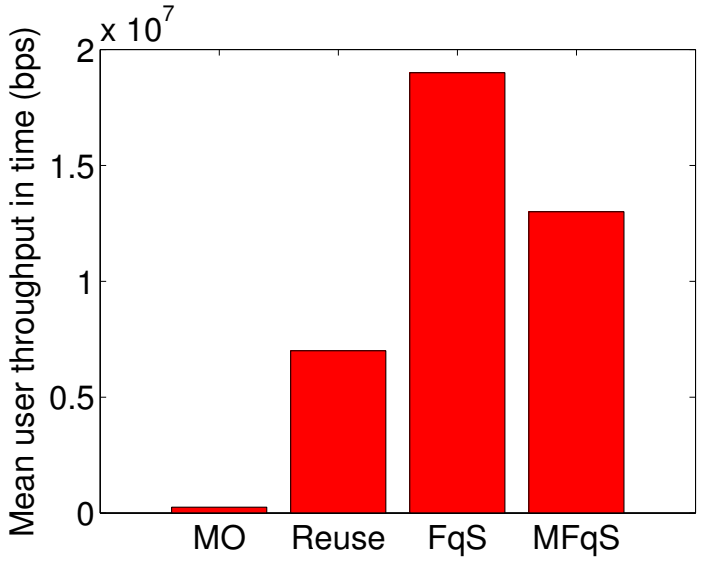


Figure 4.15: Mean user throughput in time for each bandwidth sharing scenario

been deployed in the vicinity of the traffic light, and three frequency allocation schemes have been analyzed: full frequency reuse, static and dynamic frequency splitting based on α-fair metric optimization. It is shown that due to non-stationary, periodic variations introduced by the traffic light, dynamic frequency splitting provide significant performance and QoS improvement with respect to the static one. Both frequency splitting solutions considerably reduce interference, and as a result, outperform the full frequency reuse solution.

The limited coverage zone of the small cell on the one hand and the associated cost for the backhaul infrastructure on the other hand has motivated the study of the Virtual Small Cell (VSC) technology.

In the context of Large Scale Antenna System (LSAS) technology, an heteregenous system of three antennas creating three types of focusing beams are considered in the next section: Virtual Small Cell (VSC), virtual small cell with Self-Organizing Network (VSC-SON) and Multilevel Beamforming (MB).

4.3 Focusing antenna array technology for serving users in mobility

Small cells can efficiently alleviate macro-cell loads. However this technology has two drawbaks: the cost of backhauling, and its coverage zone which is fixed to its vicinity. Mobile users that leave the small cell area may experience a deterioration of their quality of service e.g. when it is located near the cell edge. Interference problems may also occur when the small cell and macro cell share the same frequency band. Other technics are needed to not only decrease the costs but also to enhance the service quality of users leaving the small cells.

This section considers the same mobility pattern as in the previous section in the presence of a traffic light which causes a periodic congestion. We introduce now new antenna technologies with large antenna arrays: VSCs or Virtual Sectorization (ViS), and multilevel beamforming which is one of the potential solutions for Massive-MIMO systems [25]. The growing interest in large antenna array technology with beam focusing capabilities is due to the increase in spectral and energy efficiency it provides. This technology is likely to be among the main pillars of 5G networks in order to achieve the ambitious capacity targets [26].

VSC or ViS technology, ([27], [28]) aim at covering remotely a localized traffic zone, e.g. a hot spot. It can constitute a cell with a distinct cell ID. It is noted that the two terms have the same meaning and can be used interchangeably hereafter. The VSC can utilize the same frequency resources as the macro-cell (frequency-reuse 1).

The second technology is the VSC combined with SON. In this case, the VSC and the macro-cell comprise a single logical cell with a single cell ID, that dynamically share resources by means of a SON algorithm [28]. The advantage of this solution is the suppression of interference between the VSC and the macro-cell which compensates the reduced amount of resources available for each BS.

Multilevel beamforming has been proposed for backhaul [25], and more recently, for the RAN [29]. The set of multilevel beams define a codebook having a hierarchical form, from large coverage beams to more focused ones. When applied to the RAN, the concept of *beam planning* can apply, namely the form of the beams can be pre-optimized to suit the specific needs for coverage while taking into account basic geometric parameters of the cell. Multilevel beamforming can be applied to both Time Division Duplex (TDD)

and Frequency Division Duplex (FDD), and is motivated by its capability to handle large codebook size with reduced complexity associated to feedback and processing. The high level focusing capability allows to minimize the interference generated at neighboring cells.

The third technology is a new generalization of the concept of multilayer codebook: A heterogeneous antenna system on a single BS is considered. Each antenna can support a single or a set of beams. All the beams supported by the heterogeneous antenna system comprise a single codebook, denoted here as the *global codebook*. In this system, the scheduler selects the users to be served according to a utility function (e.g. Proportional Fair (PF), and then, the beams are selected according to a global codebook is selected iteratively in TDD or FDD systems. The advantage of this solution is the beam selection and resource allocation for the heterogeneous antenna system that is performed at a millisecond time scale, hence allowing to react instantaneously to traffic variations. The load generated by the traffic served using the different antennas is implicitly balanced. Furthermore, the transmissions from the different antennas do not interfere with each other.

The section analyzes the performance of the above three technologies in the context of non-stationary mobility. The analyzed scenario comprises a macrocell with fixed users, crossed by a road. A traffic light located close to the cell edge generates a non-stationary mobility pattern as described in Section 4.1. Periodic traffic cumulation occurs according to the traffic light periodicity. The car following algorithm 2.3 is applied to account for the impact of grouped mobility.

Data rate and scheduling

Denote by $R^*(x_i, t)$ the mean data rate of a user located at position x_i at time t. $R^*(x_i, t)$ depends on the considered scheduling, and is computed as the mean instantaneous data rate received during one second (see Fig.4.16). The PF scheduler is considered throughout this section, with a time step of $2ms$. A user $i^*(t)$ is selected by the scheduler if he maximizes a utility function defined as the ratio between its instantaneous and its mean data rates:

$$i^*(t) = argmax_{i \in Users} \frac{R'(x_i, t)}{R^*(x_i, t)}. \tag{4.10}$$

$R'(x_i, t) = min(4.4, 0.6log_2(1 + SINR(x_i, t)))$ is the instantaneous data rate with modified Shannon formula [30]). We recall that the time variation of SINR, $SINR(x_i, t)$, the signal to interference plus noise ratio of a user in

position x_i at time t, is given by:

$$SINR_s(x_i,t) = \frac{P_s h_s(x_i,t)}{N_0 + \sum_{k \neq s} P_k h_k(x_i,t)}. \tag{4.11}$$

s is defined as the best serving cell / beam for a user in position x_i. The sum in the denominator is taken over the interfering signals from other BSs. P_s is the transmitted power and the channel gain h_s includes the fading F and the pathloss. $R^*(x_i,t)$ is calculated as follows:

$$R^*(x_i,t) = (1-\epsilon)R^*(x_i,t-1) + \eta 1_{i=K(t)} R'(x_i,t-1) \tag{4.12}$$

with $\eta > 0$ being a small averaging parameter. It is noted that for high speed, coherence time becomes very short, and hence the PF converges to the RR scheduling.

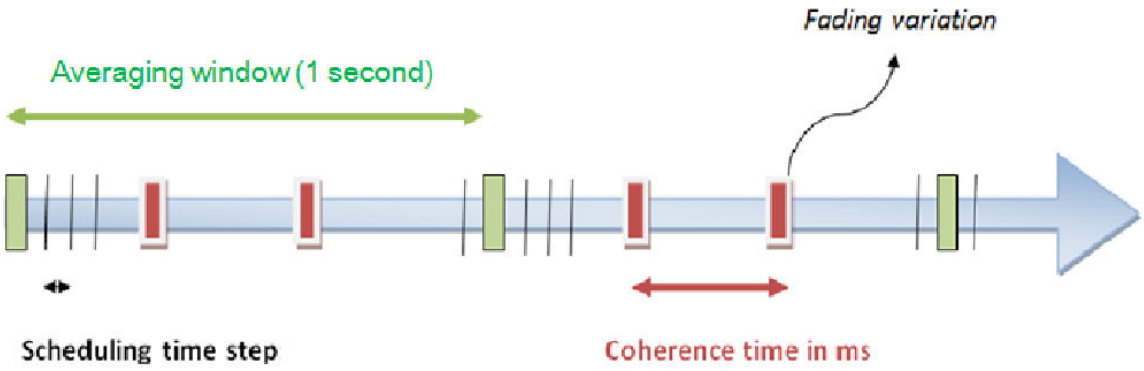


Figure 4.16: Scheduling model

We suppose in this chapter symmetric fast fading statistics, namely all the users have the same distribution of fast fading. Two different fast fading distributions are considered: *independent Rayleigh fading* and *Independent Rician fading*. The first model corresponds to an environment rich in multipaths due to high density of scatterers. The second one corresponds to an environment with significant direct, Line of sight (LoS) component between the transmitter and the receiver. Reference [31] shows that the *Rician fading* is the appropriate model in the case of opportunistic beamforming (LoS case). To simplify the analytical expression of the fading, we use the approximation of the Rician distribution by the *Nakagami model* [32] with parameters m and w:

$$f(x,m,w) = \frac{2m^2}{\Gamma(m)w^m} x^{2m-1} e^{-\frac{m}{w}x^2} 1_{x \in]0,\infty]} \tag{4.13}$$

4.3.1 Antenna array beam focusing solutions

4.3.2 Antenna Array Design

This sub-Section presents the guidelines for multilevel beamforming design. The antenna array modeling is first introduced and then the design of the beam diagrams is formulated as an optimization problem. This section of Antenna array design is an extract from [29] and is added here for sake of clarity and completeness of the sequel.

Antenna model

Consider a $N_x \times N_z$ (sub) antenna array of vertical dipoles, at a distance of $\frac{\lambda}{4}$ from a square metallic conductor, with λ being the wavelength. The full antenna array generates the highest level (narrowest) beams, whereas lower level beams correspond to smaller (rectangular) sub arrays size (see Figure 4.19). If another type of radiating element is chosen, only its radiation pattern should be modified. To simplify the model, we approximate hereafter the reflector as an infinite Perfect Electric Conductor (PEC). The N_x and N_z elements in each row and column are equally spaced with distances d_x and d_z, respectively (Figure 4.17).

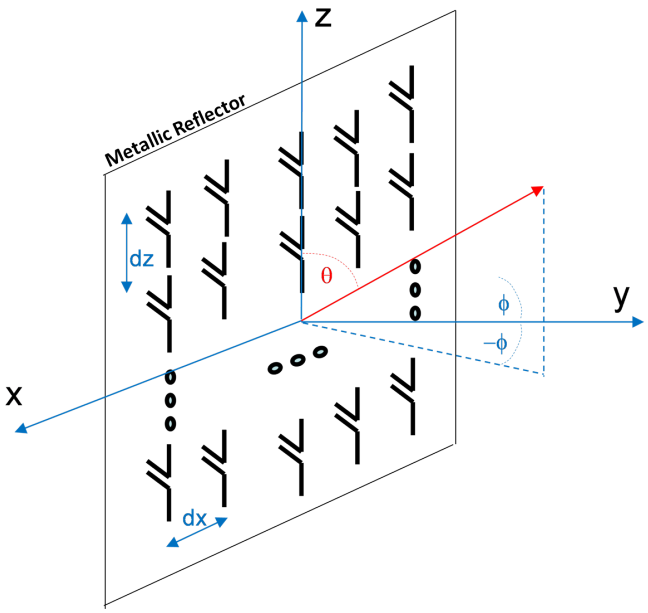


Figure 4.17: Antenna array with dipole radiating elements.

The direction of a beam is determined by the angle (θ_e, ϕ_e) in the spherical coordinates θ and ϕ. The antenna gain for a given beam defined by (θ_e, ϕ_e) in a given direction (θ, ϕ) is written as

$$G(\theta, \phi, \theta_e, \phi_e) = G_0 f(\theta, \phi, \theta_e, \phi_e) \tag{4.14}$$

where f is a normalized gain function and G_0 the maximum gain in the (θ_e, ϕ_e) direction. A separable excitation in the x and z directions is assumed, resulting in the following separable form of f:

$$\begin{aligned} f(\theta, \phi, \theta_e, \phi_e) = |AF_x(\theta, \phi, \theta_e, \phi_e) \cdot AF_y(\theta, \phi) \\ \cdot AF_z(\theta, \theta_e)|^2 \cdot G_d(\theta) \end{aligned} \tag{4.15}$$

$AF_x(\theta, \theta_e, \phi, \phi_e)$ and $AF_z(\theta, \theta_e)$ are the array factors in the x and z directions and are given by

$$AF_x(\theta, \theta_e, \phi, \phi_e) = \frac{1}{\sum_{m=1}^{N_x} w_m} \sum_{m=1}^{N_x} w_m \cdot a_m \tag{4.16}$$

and

$$AF_z(\theta, \theta_e) = \frac{1}{\sum_{n=1}^{N_z} v_n} \sum_{n=1}^{N_z} v_n \cdot b_n. \tag{4.17}$$

a_m and b_n are complex amplitude contributions of the radiating element located at $(m-1)d_x$ and $(n-1)d_z$, respectively:

$$a_m = \exp\left(-j2\pi \frac{(m-1)d_x}{\lambda}(\sin\theta \sin\phi - \sin\theta_e \sin\phi_e)\right), \tag{4.18}$$

$$b_n = \exp\left(-j2\pi \frac{(n-1)d_z}{\lambda}(\cos\theta - \cos\theta_e)\right). \tag{4.19}$$

The weights w_m and v_n for radiating elements in the m-th row and n-th columns define a Gaussian tapering function used to control the sidelobe level of the gain pattern

$$w_m = \exp\left(-\left(\frac{(m-1)d_x - \frac{L_x}{2}}{\sigma_x}\right)^2\right), \tag{4.20}$$

$$v_n = \exp\left(-\left(\frac{(n-1)d_z - \frac{L_z}{2}}{\sigma_z}\right)^2\right), \tag{4.21}$$

where L_x and L_z are the array size in the x and z directions respectively, with $L_x = (N_x - 1)d_x$ and $L_z = (N_z - 1)d_z$. The values for $\sigma_s, s = x, z$, are defined by fixing the ratio between the extreme and center dipole amplitudes respectively to a given value of α_s:

$$\sigma_s^2 = -\left(\frac{L_s}{2}\right)\frac{1}{\log(\alpha_s)}; s = x, z \tag{4.22}$$

The impact of the PEC can be modeled by replacing it with the images of the radiating elements it creates. The term $AF_y(\theta, \phi)$ takes into account the images and is written as:

$$AF_y(\theta, \phi) = \sin\left(\frac{\pi}{2}\sin\theta\cos\phi\right) \tag{4.23}$$

The normalized gain pattern of the dipoles, $G_d(\theta)$, is approximated as

$$G_d(\theta) = \sin^3\theta. \tag{4.24}$$

The term G_0 is obtained from the power conservation equation:

$$G_0 = \frac{4\pi}{\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}}\int_0^{\pi} f(\theta, \phi)\sin\theta\, d\theta d\phi}. \tag{4.25}$$

A beam is defined by the (rectangular) sub-array size, and the couple (θ_e, ϕ_e) defines its direction.

Antenna pattern optimization

The (sub) antenna array design constitutes an optimization problem with two objectives: maximizing the antenna gain (or conversely, minimizing the width of the main lobe) and minimizing the side-lobes' level, as a function of the parameters d_s and α_s, $s = x, z$. The problem is written as a constrained

optimization problem:

$$\underset{d_x,d_z,\alpha_x,\alpha_z}{\text{maximize}}\, G_0(N_{x,\max}, N_{z,max}, d_x, d_z, \theta_e, \phi_e) \tag{4.26a}$$

$$\text{s.t.} \tag{4.26b}$$

$$\max_{N_x,N_z}\{SL(N_x, N_z, d_x, d_z, \theta_e, \phi_e)\} \leq Th; \tag{4.26c}$$

$$N_{s,\min} \leq N_s \leq N_{s,\max};\ s = x, z; \tag{4.26d}$$

$$0 < \alpha_s \leq 1;\ s = x, z; \tag{4.26e}$$

$$0 < d_s \leq \frac{\lambda}{2};\ s = x, z; \tag{4.26f}$$

$$\theta_{\min} \leq \theta_e \leq \theta_{\max}; \tag{4.26g}$$

$$-\phi_{\max} \leq \phi_e \leq \phi_{\max}; \tag{4.26h}$$

where $N_{s,\min}$ and $N_{s,\max}$ are respectively the minimum and maximum number of antenna elements in the s direction, $\theta_{\min}, \theta_{\max}, \phi_{\min}, \phi_{\max}$ are respectively the minimum and maximum electrical elevation and azimuth angles of the antenna array. The constraint (4.26c) reads: the maximum side-lobe level for the whole range of sub-array sizes should be below a predefined threshold Th. The optimization can be carried out for the largest array size only. Then for discrete α_s values (e.g. with steps of 0.1), the optimization problem can be carried out exhaustively.

It is noted that for small elevation electrical tilt values, the projection of the beams on the plane is likely to spread out. Special care should be taken when setting the $\theta_{\min}, \phi_{\min}, \phi_{\max}$ angles in order to avoid overshooting on neighboring cells. These angles will depend on the geometrical characteristics of the cell (original coverage area of the considered BS before deploying the antenna array), such as the cell shape, size, and antenna height. One can consider optimizing the antenna for a wide range of elevation and azimuth angles, and then, according to the cell geometry, construct a codebook of beams for a desired angular range. Furthermore, a database with a set of codebooks can be pre-optimized for a set of cell geometries and then, according to the specific cell deployment, the most suitable codebook can be selected.

4.3.3 Multilevel beamforming for heterogeneous antenna system

We consider a heterogeneous antenna system comprising a macro-cell covering the entire cell (for the fixed users) and two antenna arrays that are able to

follow the users along the road (see Fig.4.18) moving from the left to the right of the cell.

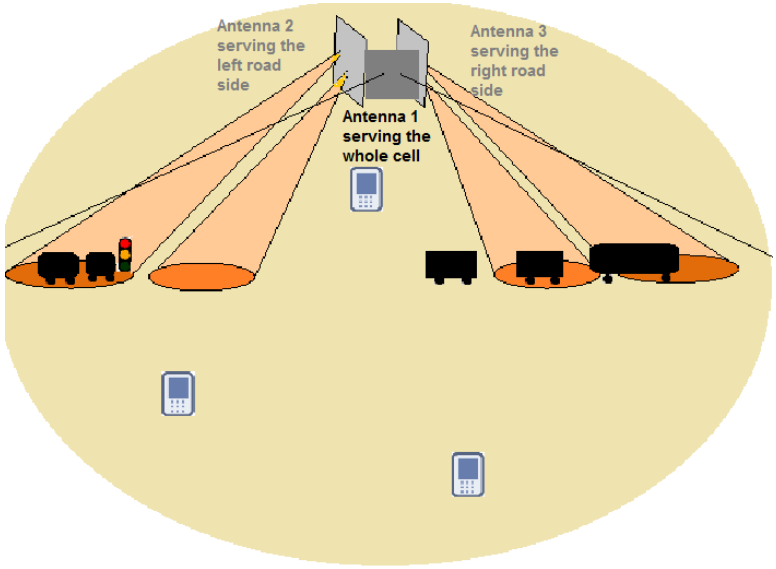


Figure 4.18: Two antennas with multilevel beamforming capability cover the traffic along a road crossing the cell

Beams structure concept

Consider a multilevel codebook as shown in Figure 4.19, with L levels and J_l beams at the level l, $l = 0, ..., L$. The jth beam at level l is written as $B_{l,j}(\theta, \theta_{l,j}, \phi, \phi_{l,j})$, $j = 1, ..., J_l$, and for brevity of notation - as $B_l(j)$. It is noted that the angles $(\theta_{l,j}, \phi_{l,j})$ correspond to (θ_e, ϕ_e) defined in section 4.3.2.

The beam of the first level, namely level 0 in Figure 4.19, $B_0(1)$, covers the entire cell. Beams at level l are generated by the same sub-array, i.e. with the same number of array elements, and differ from each other by the angles $(\theta_{l,j}, \phi_{l,j})$. Denote by $\hat{C}$ a coverage function which receives as argument a beam, and outputs its coverage area (often denoted in the cell planning nomenclature as the *best server area*), where the beam provides the strongest signal with respect to other cells or beams. The coverage function can be obtained by a network simulator for example.

The multilevel structure of the beams in the codebook means that a given beam $B_l(j)$ at level l where $l < L$ has two children beams $B_{l+1}(2j-1)$ and

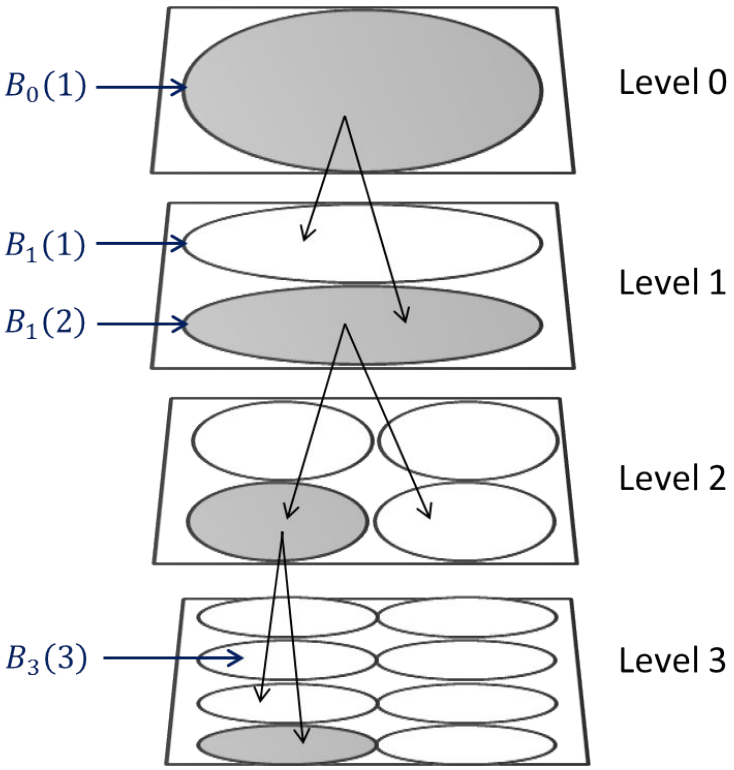


Figure 4.19: Example of beam hierarchy

$B_{l+1}(2j)$ with

$$\hat{C}(B_{l+1}(2j-1)) \cup \hat{C}(B_{l+1}(2j)) \subset \hat{C}(B_l(j)) \tag{4.27}$$

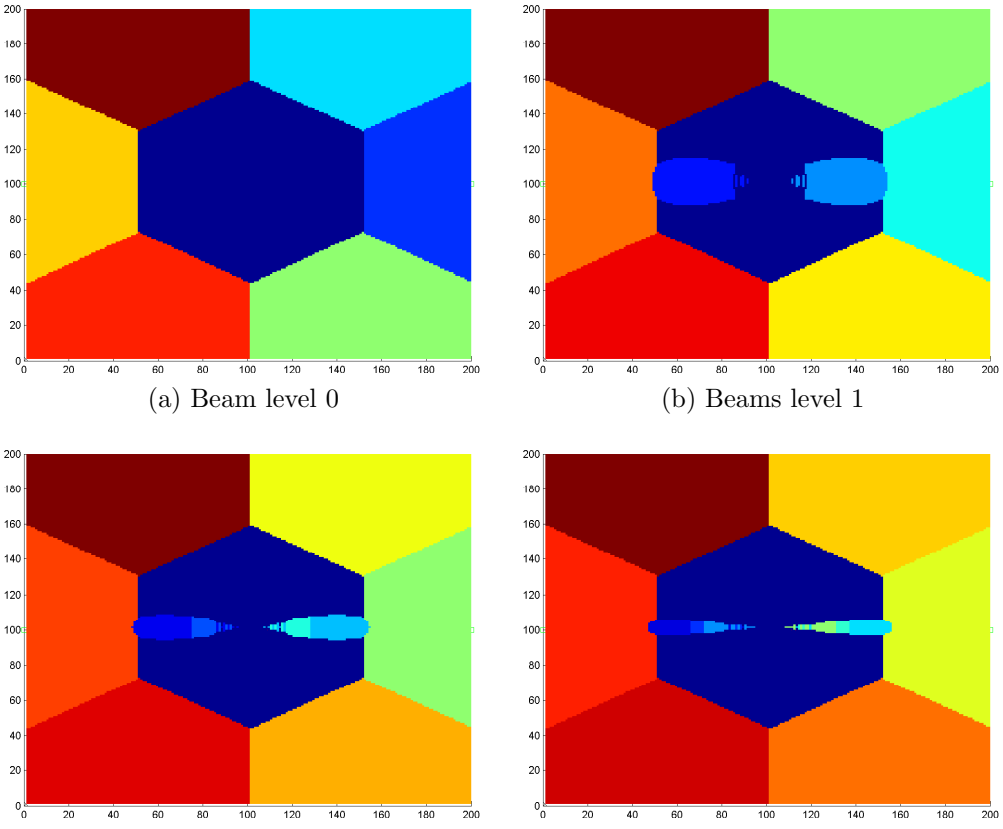
The beams at level L are the narrowest that can be obtained given the $N_{x,max} \times N_{z,max}$ antenna array.

Multilevel global codebook for the studied case

The objective of this study is to focus all the BS power on a user at any position along the road. Each antenna array utilizes a multilevel beam structure that allows to focus the transmitted signal on users on the road see Figs.4.20a-4.20d.

The beamforming with multilevel global codebook for the heterogeneous antenna system is described presently:

- Level 0 (Fig.4.20a): A global coverage from the central macro-cell antenna (Antenna 1 Fig.4.18) is used to select the user to be served;



(a) Beam level 0

(b) Beams level 1

(c) Beams level 2

(d) Beams level 3

Figure 4.20: Beamforming with global multilevel codebook adapted to the road case

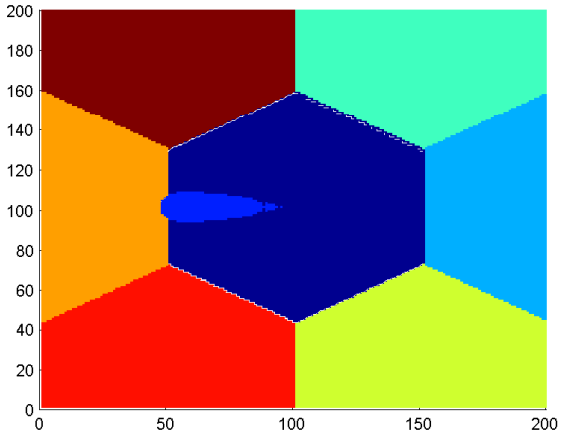
- Level 1 (Fig.4.20b): Two beams are transmitted by the two parallel antenna arrays (antennas 2 and 3) to cover users on the road on the left and right of the BS;
- Level 2 (Fig.4.20c): each beam of level 1 is split into two more focused beams to achieve higher rates for vehicular users in the road;
- Level 3 (Fig.4.20d): beams of level 1 are split into 3 more focused beams than in level 2.

To choose the user to schedule we use the PF algorithm with the best beam for each user over all the levels. To reduce complexity, one can design a scheduling algorithm in which the PF utility of a user is calculated for its present beam in level i and its direct neighbors in level $i+1$ and $i-1$ in the multilevel tree of beams. One could think that at any case a user is served with the most focused beam (i.e. at level 3). However as {Beams level i}$\not\subseteq$ {Beams level i-1}, if a user is between two beams in level 3 it is preferable to take an appropriate beam in a lower level (note that in this example, in the codebook construction condition (4.27) was relaxed).

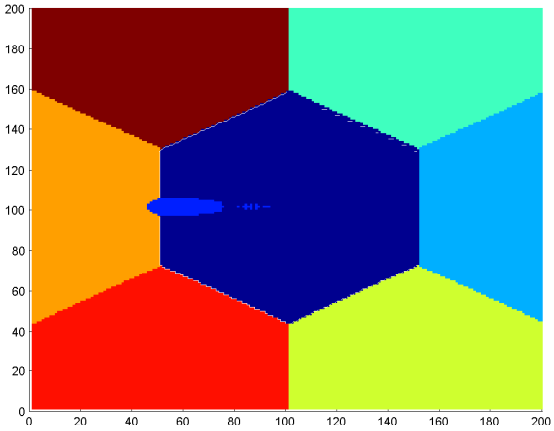
4.3.4 Virtual small cell (VSC)

The VSC is a remotely created small cell using an antenna array that radiates a fixed narrow beam. Hence the motivation is rather to cover a limited area with concentration of users rather than to follow them along their trajectories. The beam is directed towards the desired coverage area using antenna array techniques [28]. The VSC antenna can be located beside the macro-cell BS antenna, thus avoiding backhaul deployment cost as in classical small cells. The VSC beam can either operate on the same frequency bandwidth as the macro-cell (and is denoted here as *VSC*) or can dynamically share the bandwidth using a SON algorithm (and is denoted as *VSC-SON*).

While the smaller antenna which supports a wider beam in Figure 4.21a covers all the road, it is less focused and achieves lower SINR than the the more focused beam in Fig.4.21b. In both levels, one can see the antenna gain for beam level 1 in Figs.4.22 and 4.23 and for the beam level 3 in Figs.4.24 and 4.25. The latter is designed to mainly cover the area of the traffic light where periodic congestion may occur. The resource allocation for VSC and VSC-SON is described presently. The two cases of bandwidth sharing between macro and the virtual small cell are therefore analyzed: the full bandwidth reuse VSC and the optimal bandwidth sharing which takes into account users configuration at each iteration VSC-SON.



(a) $N_x \times N_z = 6 \times 16$



(b) $N_x \times N_z = 24 \times 32$

Figure 4.21: VSC antenna array with different array size $N_x \times N_z$

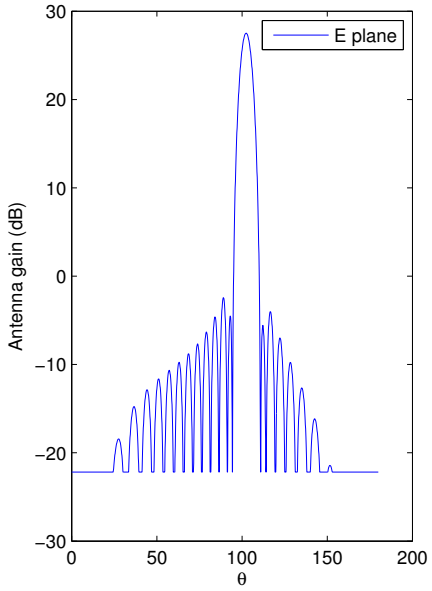


Figure 4.22: Level 1 beam diagram in θ

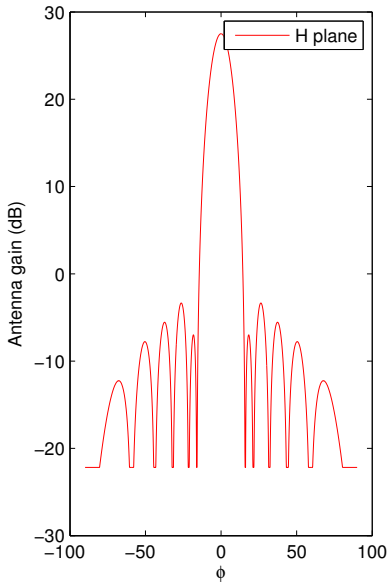


Figure 4.23: Level 1 beam diagram in ϕ

Creating a virtual small cell next to a hotspot (i.e. next to the traffic light) has a significant impact on performance indicators. For this first case we deploy a small cell next to the traffic light with a full reuse bandwidth. Full frequency reuse consists of allocating the entire bandwidth to both the macro-cell and the VSC. A fixed power budget P is assumed, namely the BS power is shared equitably between the VSC, P_v, and the macro-cell antenna, P_m, $P/2 = P_m = P_v$. In this scenario, Significant interference can be expected between the macro-cell and the VSC.

4.3.5 VSC-SON

The frequency bandwidth W is shared between the macro-cell and the VSC. The optimal bandwidth sharing takes into account the configuration of all users at each time step. The frequency bandwidth is shared dynamically between macro-cell and small-cells in their coverage area. The split factor δ is optimized at each iteration as a function of the configuration of users as previously computed in section 4.2.1.

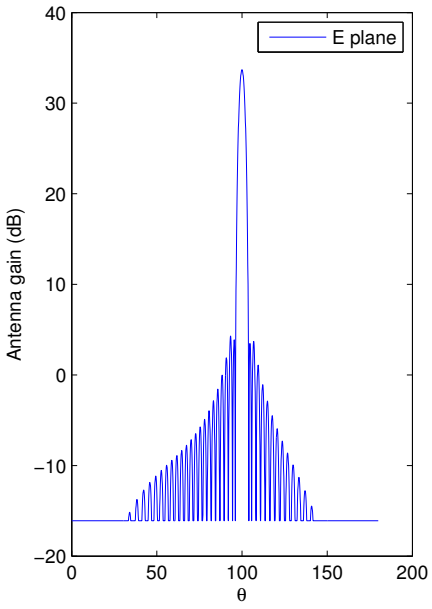


Figure 4.24: Level 3 beam diagram in θ

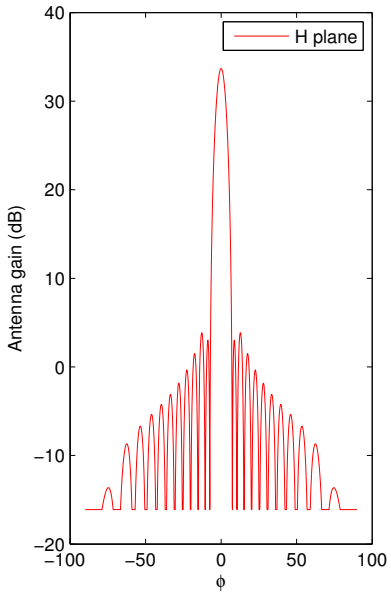


Figure 4.25: Level 3 beam diagram in ϕ

4.3.6 Simulations and numerical results

We consider the scenario 2 in Figure 4.1 of a cell with a BS in the center and a road crossing it. We want to serve users with the highest throughput possible. The traffic light is always located at the cell edge and creates the periodic congestion. Six technologies scenarios are considered: macro-cell only, four VSC cases with fixed beams equivalent to level 1 and 3 (see Figure 4.21 and Table 4.3) and finally the beamforming with mulitevel global codebook. For this section, the network and the traffic characteristics are presented in Table 4.2.

Technologies performance comparison

We compare the six scenarios cited above. The group of users cumulated by the red light will end up leaving the VSC coverage and therefore another antenna system is sought to follow the vehicular users in the cell to provision high data rates. The multilevel beamforming with the global codebook could be an effective solution to serve the moving user along the road. The performance indicators covering the different cases are depicted in Figures 4.26 for the MUT, in Figures 4.28-4.29 for the mean throughput outage probability and in Figure

Table 4.2: Network and Traffic characteristics

Network parameters	
Number of macro BSs	1
Number of interfering BSs	6
Macro-cell layout	hexagonal trisector
Intersite distance	$500m$
Bandwidth	$20MHz$
Channel characteristics	
Thermal noise	$-174dBm/Hz$
Macro Path loss (d in km)	128.1 + 37.6 $\log_{10}(d)$ dB
Data rate averaging parameter ϵ	0.2
Mobility traffic characteristics	
λ_1 mobile users	0.5 $users/s$
λ_2 fixed users	0.5 $users/s$
Position traffic light	next to the BS
Traffic light periodicity	200 it. red/ 200it. green
Maximal user speed	$5m/it.(1\ iteration = 1\ sec)$
Speed parameter a'	0.8
File size for mobile users σ_1	80 Mbits
File size for fixed users σ_2	1 Mbits

4.27 for the mean FTT. One can see that the performance of the multilevel beamforming technology outperforms the VSC technology.

Table 4.3: Characteristics of VSC beams level

Beams level	**Nx**	**Nz**	**Beam downtilt**
Level 0	0	14	10
Level 1	6	16	12
Level 2	12	16	12.5
Level 3	24	32	9.5

The first point observed in the TOP curves (Figs. 4.28 and 4.29) is that for the case of macro only, as users are periodically accumulated in the traffic light at the cell edge and their data rate condition gets worst with the increased traffic congestion. The decreasing of the departure rate of users from the network and the decreasing of the throughput induce a sharp increase in the Throughput Outage Probability (TOP) as shown in Fig.4.28. We also notice

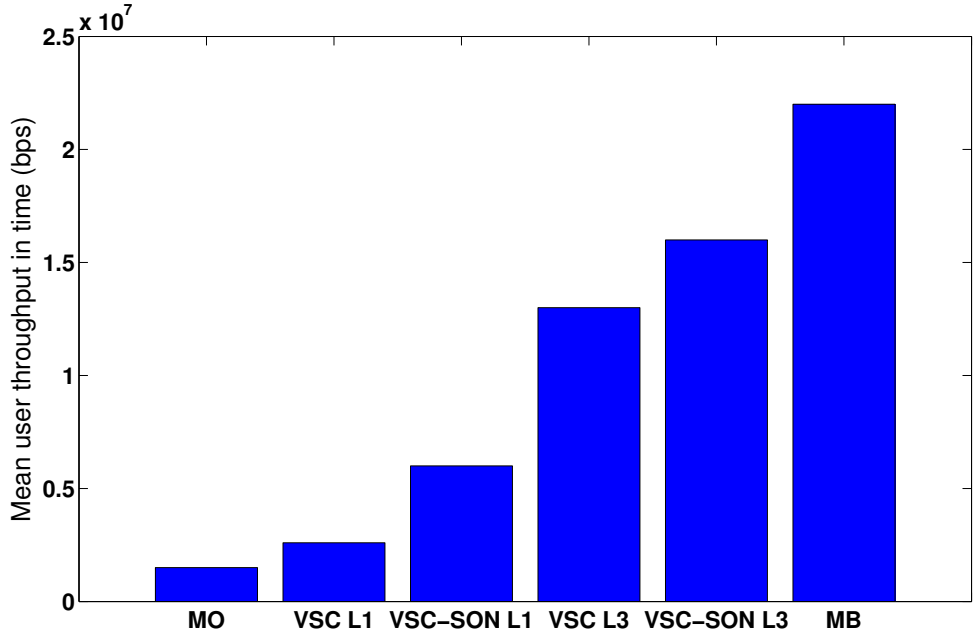


Figure 4.26: Mean user throughput in time for the six scenarios: Macro only (MO), VSC in full reuse level 1 and level 3 cases(VSC L1 and VSC L3), VSC-SON level 1 and level 3 cases (VSC-SON L1 and VSC-SON L3), multilevel beamforming (MB)

that for the 6 cases the periodic behaviour of the curves is reduced as the degree of beam focusing is increased. We summarize below the impact of the different technologies on the achieved performance:

- *VSC with beam level 1*: full reuse and bandwidth sharing are both considered. In this case, the bandwidth sharing scenario is three times better than the full reuse one (see Fig.4.26). The periodicity is almost lost in the bandwidth sharing (VSC-SON) case as departure rate of users is getting higher (better data rate is achieved). Moreover, the more the VSC beam is narrowed the higher are the cell and users' performance.

- *VSC with beam level 3*: For the case of the narrowest beam (level 3), VSC and VSC-SON achieve similar MUT but differ in the FTT. The impact of the macro-cell interference on the SINR becomes insignificant as the beam gets narrower: in Fig.4.26 MUT in VSC-SON is only 10% higher than that of the VSC MUT in level 3 comparing to three times

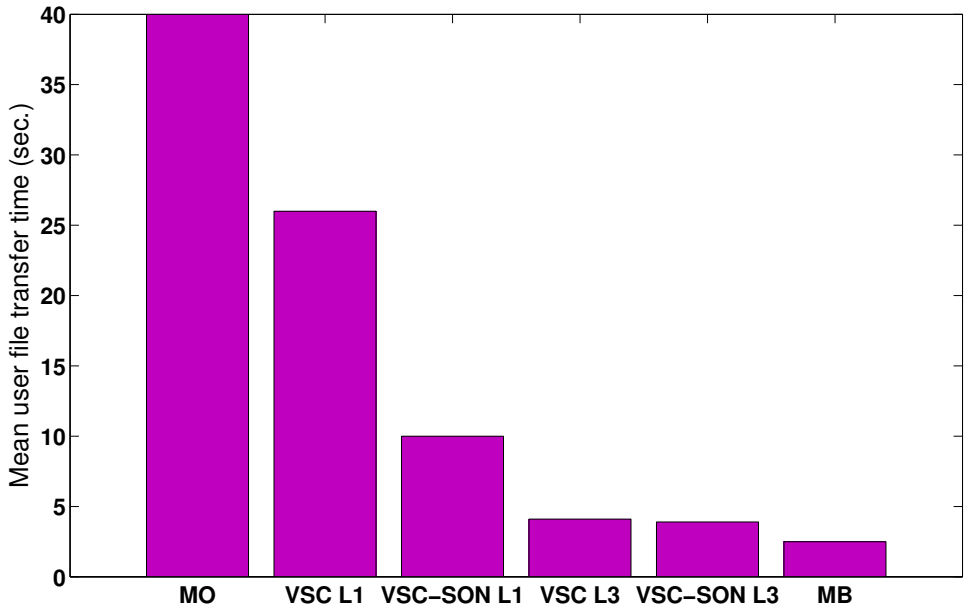


Figure 4.27: Mean file transfer time for the six scenarios: Macro only (MO), VSC in full reuse level 1 and level 3 cases(VSC L1 and VSC L3), VSC-SON level 1 and level 3 cases (VSC-SON L1 and VSC-SON L3), multilevel beamforming (MB)

at level 1.

- *Beamforming with multilevel global codebook*: one might expect that the radio performances of the cell for the multilevel beamforming would be the same as the VSC as it is the same beam focalised at the traffic light location for both cases. What makes the difference between these two cases is that vehicular users will leave the VSC as soon as the traffic light turns green. The impact of users leaving the VSC is shown through the high peaks in TOP: probability varying between 1.3% and 3% for the full reuse in beam level 3 (see Fig.4.29). These peaks are not anymore periodic in these cases as the users leave faster the network as data rate is bigger. The TOP peaks disappear in the beamforming scenario (see Fig.4.29).

  The mean TOP in this case equals 0.7%. Users who leave the VSC have a significant impact on the network performance: MUT for multilevel

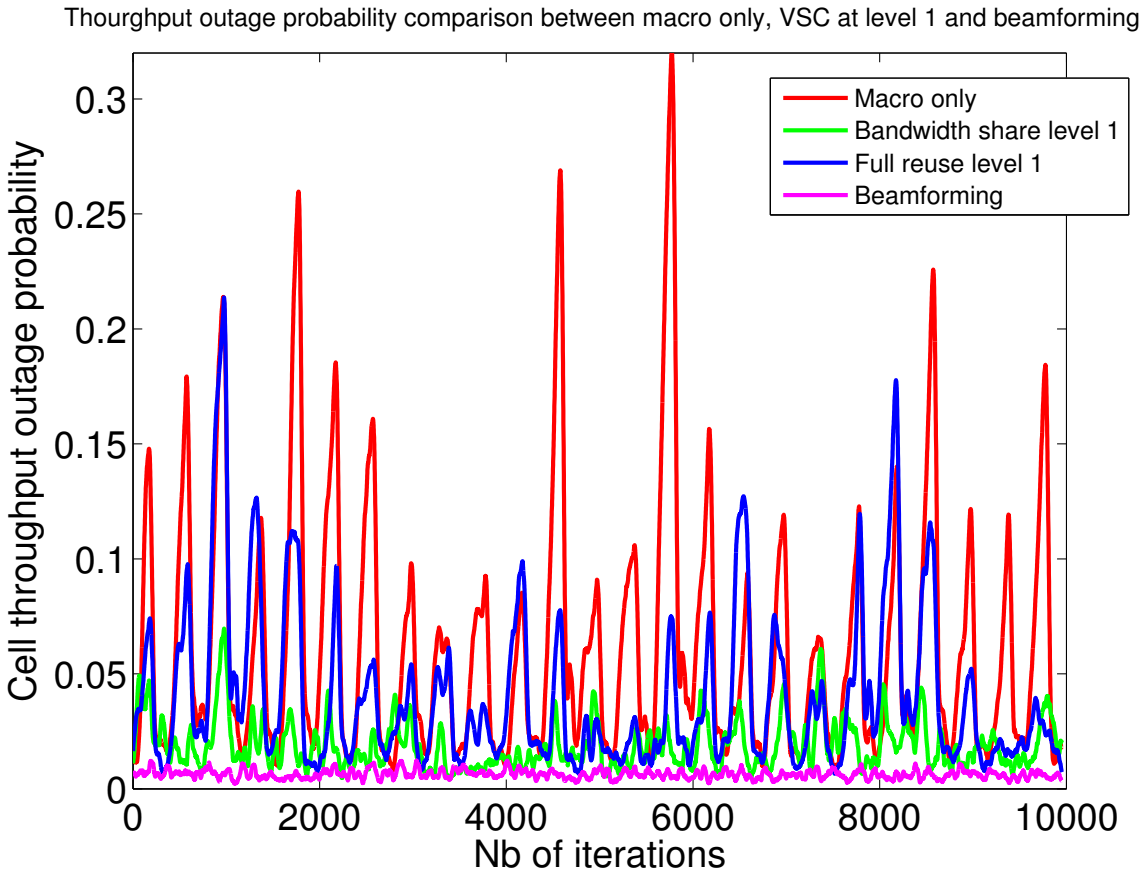


Figure 4.28: Throughput outage probability for macro-cell only (Macro only), VSC in level 1(Full reuse level 1), VSC-SON in level 1(Bandwidth share level 1) and multilevel beamforming (Beamforming) scenarios

beamforming is 40% higher than that of of the best VSC scenario i.e. with bandwidth sharing at level 3 (Fig.4.26).

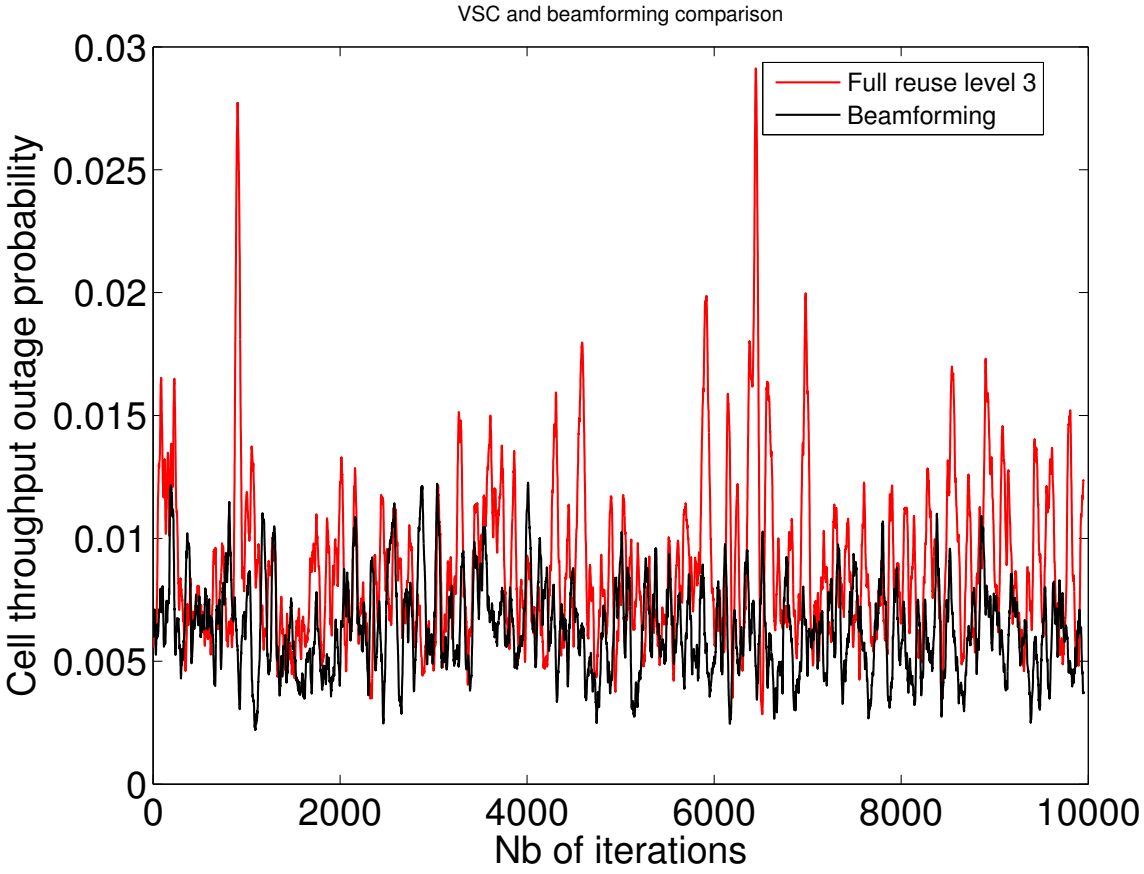


Figure 4.29: Throughput outage probability for VSC in level 3 (Full reuse level 3) and multilevel beamforming (Beamforming) scenarios

Summary

This section has shown the added value of three antenna array technologies to the performance of users with non-stationary mobility: VSC, VSC-SON and beamforming with multilevel global codebook. In the three cases, it has been shown how adapting the resource sharing and the shape of the beams to the problem that can considerably boost performance. The more focused the beams are along the road, the higher the MUT achieved. Furthermore, it has been shown how periodic outage due to traffic light periodicity and the associated traffic accumulation can be eliminated by deploying a small cell next to the physical congestion or including higher beam focusing in mobile group of users.

In static beam settings, adapting the allocated resources according to the traffic distribution using a SON algorithm can considerably increase the MUT. The best performance is achieved by the multilayer beamforming with global codebook. The reason is twofold: (i) additional focused beams are added which cover the entire road, and (ii) the ressource allocation between the three antennas is made almost at real time which implicitly balances the traffic between all the antennas.

The efficiency of opportunistic scheduling reduces with the increase of user speed due to the reduction of the coherence time. In the last chapter we develop an opportunistic scheduling model, namely the **Forecast Scheduling** that exploits high mobility and long term user diversity thanks to a priory knowledge on SINR along the mobile trajectories.

Chapter 5

Resource management for high-mobility: Forecast Scheduling

*"**RESOURCE MANAGEMENT** and service quality control in context of mobility in LTE/LTE advanced networks."*

The aim of this chapter is to fairly share the resource to all users knowing their future mobility behaviours.

We give the following example to introduce the new predictive scheduling concept: we suppose that we have two sons and we have a quantity of money to share between them. If we are fair parents, we will give half of the money for both of them. Moreover, if we are parents that make forecasts, and if we know that one of our sons will leave home tomorrow to study abroad, we will give to this son more than half of the money. Why ? Because we will not be able to give him more resource after he will leave home.

This chapter introduces a new way for resource allocation between users in high-mobility. This allocation model needs all the data information of SINR of the past, present but also the future. Based on the simple $\alpha-$fair scheduler, a generalisation of PF (equation 4.10)), the **Forecast scheduler** takes advantage of the future trajectories.

As a random event can occur at any time in the future, namely an arrival, a departure, a trajectory switch..., the Forecast Scheduling (FS) cannot predict this event. A forecast scheduling algorithm is therefore introduced taking into account such random event.

5.1 Introduction

The idea of improving resource allocation by anticipating the future state of the mobile user has recently became an active research arena, mainly for video streaming services, see for example [33],[34] and [35] which are briefly summarized below:

- In the first work ([33]), an optimal solution based on Markov Decision Processes and real content channel traces from vehicular users is provided to respect the rate requirement of the content during users' mobility. This solution is a dynamical solution for a short-term issues that allocates higher rates when channel deterioration is anticipated.

- In the second work ([34]), a piecewise linear program is given to provide an optimal solution for steady video delivery resource allocation under maximum average quality constraints for multiple users. Anticipatory networks are used to make prediction of the capacity, signal throughput data.

- In the third work ([35]), the authors propose a new qualitative scheduling mecanism for adaptive video streaming that takes advantage of future capacity in the mobile users' trajectories and show that the optimal scheduling policy is of threshold type.

This research area has been denoted as anticipatory or proactive resource allocation.

The present work considers elastic type of traffic and proposes a solution that integrates any desired degree of fairness via the $\alpha-$fair formulation. Furthermore, to our knowledge, this is the first contribution that investigates the utilization of a REM in the design of an advanced scheduler.

Thanks to the MDT technology, networks can access SINR information with GLM and be used to improve scheduling considering mobility of users. We introduce the concept of Forecast Scheduler for users in high mobility. It is assumed that a REM can provide interpolated SINR values along the user trajectories. Mobile users experience in their trajectories different mean SINR values. In mobile networks, schedulers exploit channel quality variation by giving the signal to the user experiencing best channel conditions while remaining fair. Nevertheless, we cannot record data rates of users with high mobility due to a very small coherence time. The FS will exploit the SINR variation during users' trajectories. This scheduling is formulated as a convex optimization problem namely the maximization of an $\alpha-$fair utility function

of the cumulated rates of the users along their trajectories. The solution of this problem is given by using **CVX** algorithm. However this algorithm takes a long time in the simulations. An analytical solution is introduced in this chapter for the case of 2 users and an algorithm using this solution is given for the general case. An adaptation of the algorithms to account for random events such as user arrivals or uncertainty in the selected user trajectory are developed. Numerical results for different mobility scenarios illustrate the important performance gain achievable by the FS with different algorithms.

5.2 Context

Previous schedulings exploit the channel quality variations due to several causes such as fading, shadowing etc., to select the *best user* [36]. Other schedulers share equally the throughput between users (RR). The opportunistic scheduling algorithms (α-fair scheduler for example) greatly enhance network QoS by considering the instantaneous fluctuation of the channel. However this kind of scheduling assumes that the user population is static, i.e. there is no mobility and no users' number variation. Moreover the coherence time of the channel when users drive with speed higher than $30km/h$ becomes much smaller than the scheduling delay (due to Doppler spread [37], [38]). This effect distorts the best user selection. In this case, the best way to schedule users is to use the RR scheduler. This scheduling technique remains simplistic and does not benefit from opportunistic gain related to user mobility.

If mobile users experiment different SINR along their trajectories, then their present ressource allocation could benefit from the knowledge of their future *received* data rate.

In order to explain the idea, we introduce a simple example: We suppose that there are two users in a cell: a user 1 fixed and a user 2 mobile. These two users are in the same radio condition at time $t = 1$, i.e. they receive the same data rate. As the user 1 stays at time $t = 2$ in the same location, his mean throughput remains fixed while the user 2 has moved at time $t = 2$ to a worse (or better) data rate (see Fig.5.1). Using the proportional-fair (or any instantaneous opportunistic scheduler) in this case is no longer fair for both users. This scheduler gives half of the resource to user 1 and half to the other at time $t = 1$ and will give all the resource at time $t = 2$ to the user with best radio conditions. If we had known that the user 2 will be at $t = 2$ in worse (resp. better) condition than that at time $t = 1$ then we would give him more (resp. less) than half of resources at time $t = 1$.

The aim of this study is to find a model to allocate resources fairly to all

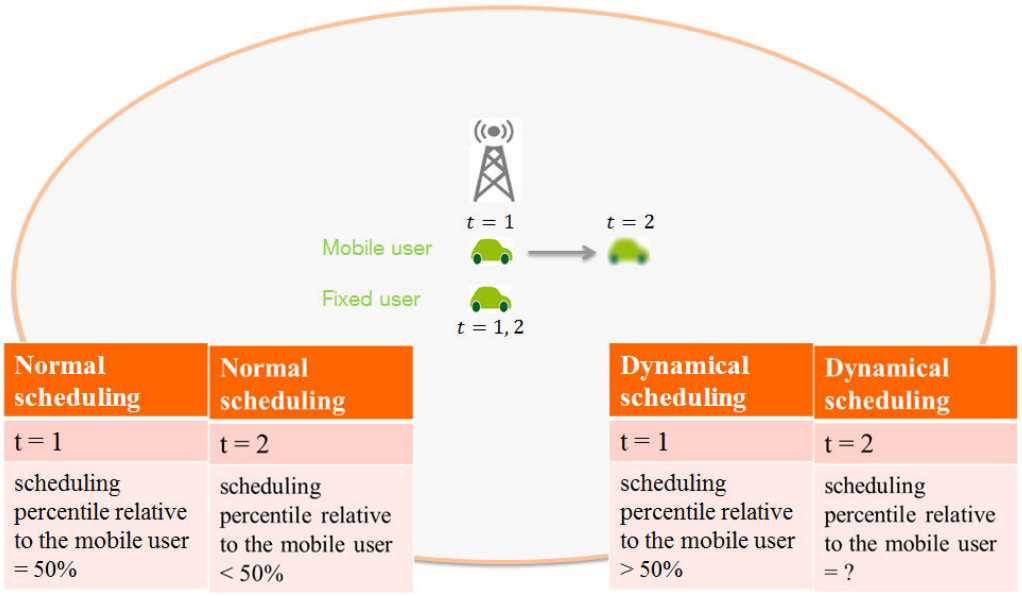


Figure 5.1: Comparison of Forecast Scheduler and alpha-fair scheduler in the context of two users in two time instances

the network users (fixed and mobile at the same time) having data prediction (SINR, received power...), users' speed and users' trajectories.

5.3 Data collect and prediction

As fast moving vehicular users change continuously their locations, they will experience different data rates depending on their future trajectories. We want therefore to know their predictive trajectories and speed to compute the position reached at any time in the future, and the possible data rates achieved during their travels (see Fig.5.2 where $\phi(S_i^t)$ is the data rate achieved by the user i at time t and is a function of the SINR S_i^t).

We propose in this section different ways to provide/collect/compute data in each user's trajectory to use for computing the predictive scheduling model.

5.3.1 SINR /received power prediction

To collect the received power at any location, in order to compute users' $SINR$ during their trajectories, two ways are proposed in this study: Drive-test and

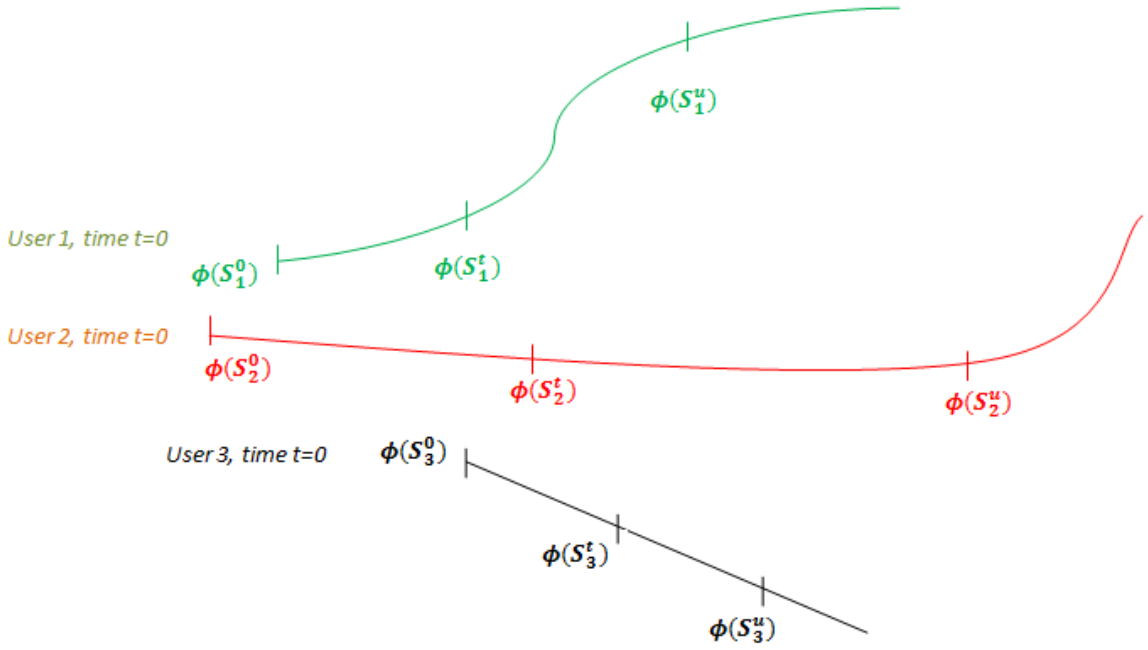


Figure 5.2: Example of 3 users with 3 different predictable trajectories. $\phi(S_i^t)$ is the predicted data rate of user i at time t.

MDT:

Drive-test:

The drive testing consists on gauging the QoS received by measuring SINR or received power. This technic is usefull in our study since we need the SINR prediction during the users' travel. However the signal is very noisy and we cannot detect a fixed power over time. One of the solutions given is to ignore the uncertainty of the future. We compute the mean value of the signal power depending on the historical observations in a particular location [39]. The main idea of this solution relies on the intuition that *Past Tells More Than Present*. In fact if the signal randomness is independant in time, it is preferable to take directly the mean.

To compute the mean of the signal power with a variance less that 10%, a sample of at least 75 items (in time) of signal power value is needed ([39]). However as the traffic studied in this work is supposed dynamic, we need to measure the data in different times periods: we can either compute the mean at each hour by using 75 data points, or at each two hours, etc. In the extreme

case, to have a dynamical variation of the received power, one can measure the signal at the same hour at each second during 75 days and average over this period. The drive testing may prove to be laborious in this case and another method can be used: *MDT data collection*.

MDT data collection:

The use of GLM for the optimization and troubleshooting of the RAN is one of the challenging topics in future RANs. The feature that allows to perform such measurements has been introduced into the LTE standard under the term MDT. The term MDT was motivated by the need to replace costly drive tests needed to manage and to troubleshoot the network by mobile based automatic GLM. However, as is shown in this chapter, GLM has the potential to be used not just for troubleshooting the RAN but for designing powerful Radio Resource Management (RRM) algorithms that can considerably improve the network performance.

MDT measurements can be performed in *immediate* or *idle* mode [40]. In immediate mode a connected mobile performs the GLM and immediately reports them to the BS. In idle mode the mobile performs the measurements according to the predefined configuration and reports them to the BS once it enters a connected mode. The MDT feature is presently available in mobile chipsets but not yet activated.

The perspective of having GLM has opened an active research and development domain, namely the construction of a REM using spatial interpolation techniques ([41], [42]). The REM can be created and updated in a MDT server in the management plane and be downloaded into each BS (see Fig.5.3). It can provide maps for different quantities such as the received signal strength or SINR. The BS can then use the REM to optimize RRM algorithms e.g. for association, handover, or resource allocation.

5.3.2 Coverage prediction

In general we cannot measure the signal power continuously in all locations. We can (for example) interpolate the available observations to have a continuous coverage description. But this technique remains simplistic and unrealistic especially when the observed measurements are random due to shadowing and the unknowledge of the area. The paper [41] suggests to predict the coverage using the Kriging spatial prediction technique, and more precisely Fixed Rank Kriging (FRK) [43] to reduce the complexity of the algorithm.

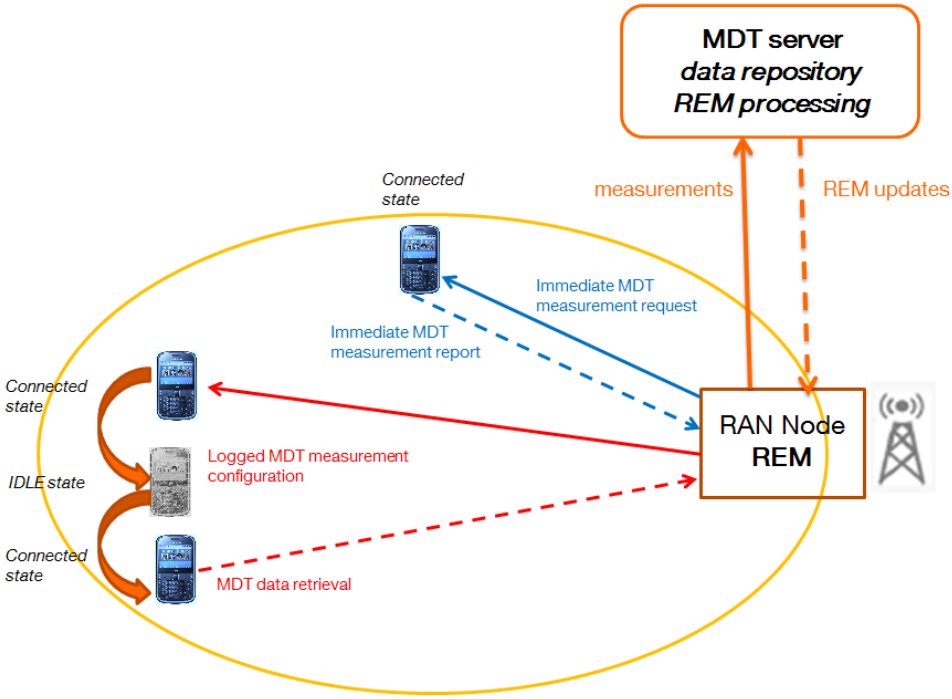


Figure 5.3: MDT data collection and coverage prediction

Another approach is to use the ray tracing based on physical optics, physical theory of diffraction and fields related to electromagnetics [44].

5.3.3 User speed prediction

Vehicle speed prediction becomes a very important topic in different traffic management fields. Knowing the future fluctuation of vehicle speed remains important for energy management to achieve optimal fuel economy for instance. In this section we suppose that the speed prediction provides the future user location and therefore the signal power in these locations. Traffic information, either historic or real time data, can be available thanks to Intelligent Transportation Systems (ITS). The traffic data is then collected using Global Positioning System (GPS), Geographical Information Systems (GIS),Vehicle-to-Vehicle (V2V), Vehicle-to-Infrastructure (V2I) [45]. The traffic information is the outcome of the aggregate traffic variables such as traffic flow, vehicular density, travel speed, travel time for a given trajectory. The traffic data process to predict vehicles speed using four different prediction models:

- ARIMA model [46],[47]: removes noise and exploits traffic conditions over a long past period;
- Non-parametric regression [48]: approaches the speed prediction by identifying the most similar past cluster;
- Kalman filtering model [49], [50]: approaches the predictive system continuously by a linear dynamic model. A mean of users' speed is estimated by minimizing the distance to the value with squarred error.

5.3.4 Trajectory prediction

In general, in public transport (bus, tramway, train...), the users trajectories are well known. these trajectories are deterministic. Moreover, users at high speeds (more than 90 km/h in their private vehicles), have also a deterministic trajectory as they cannot turn neither left nor right with such speeds. Besides these speeds are only allowed in highways and trunk roads which are deterministic trajectories. In urban areas, users speed do not exceed 50 km/h however the trajectory is unpredictable as they can turn at each time in crossroads for example. Previous works ([51]) developed trajectories prediction in order to manage the handoff and rerouting of connections problems. The study consists on developing a hierchical user mobility that imitates the user movement and use predictive algorithms (Kalman filtering techniques) for future locations. Another way to give the user future trajectory is to upload the GPS information of each user. In fact, almost all users own GPS in their private vehicles or use their phone mobile GPS. To access a location in a urban or a rural area, user needs the car or the mobile phone GPS.

5.4 System model

Consider an omni-directional macro-cell (BS) surrounded by six interfering neighbouring macro-cells. A REM is deployed in the BS and provides SINR values corresponding to the mobile location. Vehicular mobile users with a fixed speed of 50 km/h are considered along one or two trajectories. The spatial resolution of the REM with interpolation is of 1 m (it is recalled that the REM interpolates GLM), and in 50 km/h it corresponds to a 70 ms time intervals over which the SINR is considered constant. Hence the time resolution of the FS is of 70 ms. During this time interval, *a fixed allocation is applied*, namely

the same users are scheduled at a time interval depending on the technology (e.g. 1ms for LTE).

The data rate is a function ϕ of the SINR values and is given by the Shannon formula:

$$\phi(SINR) = Wlog_2(1 + SINR), \tag{5.1}$$

where W is the bandwidth.

We suppose that there are no arrivals or departures of users during the scheduling duration. Full buffer users are assumed, namely having an infinite volume to transmit. The generalization of the model to account for arrivals and departures is addressed in the Section 5.8.

5.5 Forecast Scheduling model

5.5.1 Model definition

Consider n 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: $t \in \{1, 2, .., T\} = [|1, T|]$ and i denotes the user number, $i \in \{1, 2, .., n\} = [|1, n|]$. During a scheduling period denoted here for simplicity as a unit time (e.g. 70 ms), the bandwidth is shared among the scheduled users. Let a_i denote the bandwidth proportion allocated to a user i, $a_i(t) \in [0, 1]$, according to the scheduling strategy, and W - the total bandwidth. The throughput of user i is defined by

$$a_i\phi(SINR_i) = a_iWlog_2(1 + SINR_i), \tag{5.2}$$

Denote by $\phi(S_i^t)$ the throughput of user i at time t and by S_i^t the predicted SINR. Denote by $a_i(t)$ the scheduling allocation namely bandwidth proportion of the user i at time t where $\forall t$, $\sum_{i=1}^{n} a_i(t) = 1$.

The forecast scheduling allocation policy is defined by the following optimization problem, with $\alpha \neq 1$:

$$\begin{aligned} maximize : f(a) = \sum_{i=1}^{n} \frac{(\sum_{t=1}^{T} a_i(t)\phi(S_i^t))^{1-\alpha}}{1-\alpha} \\ with : \forall i, \forall t, a_i(t) \geq 0 \\ and : \forall t, \sum_{i=1}^{n} a_i(t) = 1 \end{aligned} \tag{5.3}$$

and for $\alpha \to 1$, the optimization problem with the same constraints reads:

$$maximize : f(a) = \sum_{i=1}^{n} log(\sum_{t=1}^{T} a_i(t)\phi(S_i^t)). \tag{5.4}$$

Both problems (5.3) and (5.4) have a concave function f for $\alpha \geq 0$, and can be solved using the framework of convex optimization. For example one can use convex optimization solvers such as CVX. The size of the optimization problem is defined by the number of unknown variables, namely $n \times T$.

The formula (5.4) is used in this study. The interpretation of (5.4) is the following: resources are shared fairly among the users according to the data-rate variation in their future trajectories. For example, if a user has a coverage hole in his future trajectory, the FS will take this into account and allocate this user as much data as possible before reaching the coverage hole while remaining fair with respect to the other users. Detailed examples are given in section 5.10.

In case the number of users changes during the interval T due to a new arrival or departure in the cell, a solution is given in section 5.9.2.

The few milliseconds of the scheduling time delay gives rise to the problem of choosing the resolution method. Any approximation of this problem solution needs more than some milliseconds to be computed (Newton, Gradiant approximation, Convex optimization algorithm (CVX), etc.) which does not suit to the scheduling time delay requirements especially for high speeds. A method of resolution is provided in this work based on Karush-Kuhn-Tucker (KKT).

5.5.2 Karush-Kuhn-Tucker Resolution

The optimization problem (5.3) considered in this work is quite difficult to solve as there is a high number of unknown variables. We could consider this problem as a maximization problem with a Markovian Decision Process solution. It will be interesting to explore this way if we suppose that there is some randomness of the SINR maps. However it will take a significant time to solve the optimization problem comparing to the scheduling delay (1 to 2 ms).

As we supposed that every optimization problem parameters, namely users' SINR and their data rates, their speeds and their trajectories are known, a solution method can be considered: the KKT resolution.

We derive presently the KKT conditions for the problem (5.3). The optimization problem (5.3) has a nonlinear objective function f with regular equality and inequality conditions, i.e. differentiable constraint functions. If

the objective and constraint functions (5.3) are continuously differentiable at any $a = (a_1(1), a_1(2), ...a_1(T), a_2(1), ..., a_n(T)) \in \mathbb{R}^{nT}$, then there exist multipliers $\lambda_{k,j}$ and ν_j, where $k \in [|1, n|]$ and $j \in [|1, T|]$, called KKT multipliers ([52], Chap.5) with the following Lagrangian function:

$$L(a, \nu, \lambda) = f(a) + \sum_{j=1}^{T} \nu_j (\sum_{k=1}^{n} a_k(j) - 1) + \sum_{k,j} \lambda_{k,j} a_k(j), \tag{5.5}$$

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

We define the Lagrange dual function as the maximum value of L over a. Let a^* maximizes the Lagrangian function (5.5) for the optimal multipliers $\lambda^*_{k,j}$ and ν^*_j, where $k \in [|1, n|]$ and $j \in [|1, T|]$, therefore its gradiant is null at this point:

$$\nabla L(a^*, \nu^*, \lambda^*) = 0. \tag{5.6}$$

Which implies for all $i \in [|1, n|]$ and $t \in [|1, T|]$ the following KKT conditions:

$$\begin{aligned} \frac{\partial f(a)}{\partial a_i(t)} + \frac{\partial}{\partial a_i(t)} \sum_{j=1}^{T} \nu_j (\sum_{k=1}^{n} a_k(j) - 1) + & \\ \frac{\partial}{\partial a_i(t)} \sum_{k,j} \lambda_{k,j} a_k(j) &= 0, \\ a_i(t) &\geq 0, \\ \sum_{i=1}^{n} a_i(t) &= 1, \\ \lambda_{i,t} a_i(t) &= 0, \\ \lambda_{i,t} &\geq 0, \end{aligned}$$

which is equivalent to :

$$\phi(S_i^t)(\sum_{j=1}^{T} a_i^*(j)\phi(S_i^j))^{-\alpha} + \nu_t^* + \lambda_{i,t}^* = 0, \tag{5.7}$$

$$a_i^*(t) \geq 0, \tag{5.8}$$

$$\sum_{k=1}^{n} a_k^*(t) = 1, \tag{5.9}$$

$$\lambda_{i,t}^* a_i^*(t) = 0 \tag{5.10}$$

$$\lambda_{i,t}^* \geq 0. \tag{5.11}$$

We note from (5.7) that at any time t, for all users i and w with $w \neq i$

$$\begin{aligned}
&\phi(S_i^t)(\sum_{j=1}^{T} a_i^*(j)\phi(S_i^j))^{-\alpha} + \lambda_{i,t}^* \\
&= \phi(S_w^t)(\sum_{j=1}^{T} a_w^*(j)\phi(S_w^j))^{-\alpha} + \lambda_{w,t}^* \qquad (5.12)
\end{aligned}$$

and also from (5.7), for all user i, we have for all times t and u:

$$\frac{\lambda_{i,t}^* + \nu_t^*}{\phi(S_i^t)} = \frac{\lambda_{i,u}^* + \nu_u^*}{\phi(S_i^u)} \qquad (5.13)$$

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. The equality (5.12) explicits the resource balancing among users at each time relatively to α in the sense of equalizing the two expressions of each user.

We deduce from equality (5.12) that the choice of α impacts the user selection:

- If $\alpha = 0$ then the equality (5.12) is equivalent to $\phi(S_i^t)+\lambda_{i,t}^* = \phi(S_w^t)+\lambda_{w,t}^*$ which means that the user that will be scheduled at a future time t is the user who will reach the highest data rate at time t due to equation (5.10). In fact this particular case is the same as the normal α-fair scheduling if $\alpha = 0$ since the maximum of the sum is the sum of the maximum in this case.
- If $\alpha \longrightarrow 1$ the forecast scheduler is then called the *forecast proportional fair* scheduler.
- If $\alpha \longrightarrow \infty$ it is called the *max-min forecast fairness scheduler*.

If we suppose that the scheduler allocates at most one user (user i) at a time t then:

$$\begin{aligned}
a_i^*(t) &= 1 \\
\forall k \neq i, a_k^*(t) &= 0
\end{aligned}$$

then by (5.10) we have $\lambda_{i,t}^* = 0$. Hence:

$$\frac{\phi(S_i^t)}{(\sum_{j=1}^{T} a_i^*(j)\phi(S_i^j))^{\alpha}} = \frac{\phi(S_w^t)}{(\sum_{j=1}^{T} a_w^*(j)\phi(S_w^j))^{\alpha}} + \lambda_{w,t}^* \qquad (5.14)$$

This last equality reminds the α-fair scheduling condition in the stationnary case as we have $\lambda_{k,t} \geq 0$:

$$\frac{\phi(S_i^t)}{(\sum_{j=1}^{T} a_i^*(j)\phi(S_i^j))^\alpha} \geq \frac{\phi(S_w^t)}{(\sum_{j=1}^{T} a_w^*(j)\phi(S_w^j))^\alpha},$$

which means that at time t the user scheduled $i(t)$ is defined as following:

$$i(t) = \arg\max_k \frac{\phi(S_k^t)}{(\sum_{j=1}^{T} a_k^*(j)\phi(S_k^j))^\alpha}. \tag{5.15}$$

The system of equations corresponding to the KKT conditions is difficult to be solved analytically directly. With $n \times T + n \times T + T$ unknown variables in the system, including all the $a_i^*(t)$, $\lambda_{i,t}^*$ and ν_t^*, and just $n \times T + T$ equations, one has to try solutions and verify the if these solutions verify the optimal conditions. In the next sub-Section we show how to derive an analytical solution to the scheduling problem (5.3) for the case of two users.

5.5.3 Closed form solution for n=2

In all the calculations that follow, we rewrite $a = a^*$, $\lambda = \lambda^*$ and $\nu = \nu^*$ for sake of simplicity.

Suppose there are two users in the cell, i.e. $n = 2$ in the problem (5.3). There are still more unknown variables than equations. We introduce the following property to allow the solution of the problem:

Problem assumptions

Given certain condition on the data rates (see Theorem below), if the two (all, for the general case) users are scheduled simultaneously at some time $t = K$ then only one user is scheduled in all other times within $[1, T]$

Theorem 1. ***If*** *the following conditions are verified:*
1/ There exists a time $t = K$ where $\forall i, a_i(K) > 0$ i.e. both two users are scheduled simultaneously at time K,
2/ $\forall i, j$, $\frac{\phi(S_i^t)}{\phi(S_i^k)} \neq \frac{\phi(S_j^t)}{\phi(S_j^k)}$,
Then *$\forall t \neq K$ in $[1, T]$, there exists only one $j(t)$ such as $a_{j(t)}(t) > 0$.*

See the proof in the annexe.

From Theorem 1, one can deduce that only two cases exist:

- both users are scheduled simultaneously at time K and in other times only one of them is scheduled;
- only one user is scheduled at each time, i.e. no such time K exists.

The assumption 2 of Theorem 1 comes from the fact that the data rates fractions are almost surely not equal between users in reality as they have almost never the same trajectories at the same time.

It is noted that numerical experiments have shown that the case described in Theorem 1 does not occur often. In the next two subsections we check the existence of such time K.

For sake of brevity, we denote the sentence '*the existance of time K for which the condition 1 of Theorem 1 is verified*' as '*the existance of time K*'.

If time K exists then users are scheduled at the same time K

We suppose in the following that time K exists and that at a given time $t \neq K$ the user 1 is scheduled (and hence user 2 is not scheduled) then from equation (5.13):

$$\lambda_{1,t} = 0, \lambda_{1,K} = 0 \Leftrightarrow \frac{\nu_K}{\phi(S_1^K)} = \frac{\nu_t}{\phi(S_1^t)} \tag{5.16}$$

$$\lambda_{2,t} \geq 0, \lambda_{2,K} = 0 \Leftrightarrow \frac{\nu_K}{\phi(S_2^K)} \geq \frac{\nu_t}{\phi(S_2^t)} \tag{5.17}$$

Denote by $\psi_i^t = \frac{\phi(S_i^t)}{\phi(S_i^K)}$, then:

$$\psi_1^t > \psi_2^t, \tag{5.18}$$

The inequality is strict as assumed in 1/ of Theorem 1. User 1 is scheduled at t with the condition of existence of a time K if and only if the three conditions below are verified:

$$Inequality(5.18), \tag{5.19}$$

$$\begin{aligned} a_1(t) + a_2(t) &= 1, \\ a_1(t), a_2(t) &\geq 0. \end{aligned}$$

Theorem 2. *If we suppose that a time K exists then for all time $t \neq K$ the users 1 and 2 are scheduled with the following scheduling strategy:*

$$\begin{aligned} a_1(t) &= 1_{\psi_1^t \geq \psi_2^t} \\ a_2(t) &= 1_{\psi_2^t \geq \psi_1^t} \end{aligned} \tag{5.20}$$

We can now check if a time K exists by verifying if for both users $a_i(K) > 0$ using Theorem 2. If it is not the case then K does not exist. $a_1(K) > 0$ and $a_2(K) > 0$ can be written as following (see within proof of Theorem 3 in the Annexe):

$$a_1(K) = \frac{1}{\phi(S_1^K)}((\frac{-\nu_K}{\phi(S_1^K)})^{-1/\alpha} - \sum_{t\neq K} \phi(S_1^t)1_{\psi_1^t \geq \psi_2^t}) > 0 \tag{5.21}$$

$$a_2(K) = \frac{1}{\phi(S_2^K)}((\frac{-\nu_K}{\phi(S_2^K)})^{-1/\alpha} - \sum_{t\neq K} \phi(S_2^t)1_{\psi_2^t \geq \psi_1^t}) > 0, \tag{5.22}$$

where $\nu_K = -\frac{(\phi(S_1^K)^{1/\alpha-1}+\phi(S_2^K)^{1/\alpha-1})^\alpha}{(1+\sum_{t\neq K}\psi_1^t 1_{\psi_1^t\geq\psi_2^t}+\psi_2^t 1_{\psi_2^t\geq\psi_1^t})^\alpha}$.

It is recalled that $a_2(t) = 1 - a_1(t)$.

Using $\phi(S_i^K)\psi_i^t = \phi(S_i^t)$, denoting by $\phi_1 = \phi(S_1^K)$ and $\phi_2 = \phi(S_2^K)$, and using (5.21) and (5.22), we can summarize the result in *Theorem* 3 below.

Theorem 3. *K exists if and only if the following two inequalities are verified:*

$$\begin{aligned}
&\phi_2^{\frac{1}{\alpha}} \sum_{t\neq K} \phi(S_1^t)1_{\psi_1^t>\psi_2^t} - \phi_1^{\frac{1}{\alpha}} \sum_{t\neq K} \phi(S_2^t)1_{\psi_2^t>\psi_1^t} < \phi_1^{\frac{1}{\alpha}}\phi_2, \\
&-\phi_2^{\frac{1}{\alpha}} \sum_{t\neq K} \phi(S_1^t)1_{\psi_1^t>\psi_2^t} + \phi_1^{\frac{1}{\alpha}} \sum_{t\neq K} \phi(S_2^t)1_{\psi_2^t>\psi_1^t} < \phi_2^{\frac{1}{\alpha}}\phi_1.
\end{aligned} \tag{5.23}$$

If one of the conditions of Theorem 3 is not verified then time K does not exist and another scheduling strategy (rule) should be derived as described presently.

If time K does not exist

We suppose in this section that at time 1 user 1 is scheduled and therefore $\lambda_{1,1} = 0$. We want to verify the assumption on user 1 selection at time 1. For all time t, the sign of $\lambda_{1,t} - \lambda_{2,t}$ determines the scheduling decision: if it is positive then the user 2 is scheduled at time t ($\lambda_{1,t} \geq \lambda_{2,t} \iff a_2(t) \geq a_1(t)$ by equation (5.11)). One of the lambdas must be null and the other one positive for the case of two users.

With equation (5.13), as $\lambda_{1,1} = 0$, we have:

$$\lambda_{1,t} - \lambda_{2,t} = \nu_1\psi_1^{t/1} - \nu_1\psi_2^{t/1} - \lambda_{2,1}\psi_2^{t/1} \tag{5.24}$$

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

$$\lambda_{1,t} - \lambda_{2,t} = \nu_1(\psi_1^{t/1} - \psi_2^{t/1}) - \lambda_{2,1}\psi_2^{t/1} \tag{5.25}$$

As $\lambda_{2,1}\psi_2^{t/1} > 0$ we divide the equation (5.25) by $\lambda_{2,1}\psi_2^{t/1}$ and hence the sign studied is the same as the sign of $\frac{\frac{\psi_2^{t/1} - \psi_1^{t/1}}{\psi_2^{t/1}}}{-\frac{\lambda_{1,2}}{\nu_1}} - 1$.

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.

Theorem 4. *If user 1 is supposed scheduled at time $t = 1$, then:*
$\psi_2^{t/1} < \psi_1^{t/1} \Longrightarrow$ *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 4, i.e. if $t \in A_1$ 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 (with the corresponding proof in the Annexe):

Theorem 5. *User 1 is scheduled at time u if and only if:*

$$(\sum_{A_2^u} \phi(S_2^j))^\alpha > \max_{k \in A_1^u} \frac{\phi(S_2^k)}{\phi(S_1^k)} (\sum_{A_1^u} \phi(S_1^j))^\alpha \tag{5.26}$$

Equivalent scheduling rule can be written for user 2.

5.6 Heuristic solution for $n > 2$

In this Section, we will use the analytical result for the case $n = 2$ from the previous section to derive a forecast scheduling heuristic algorithm for any number of users. We assume hereafter that $n > 2$.

5.6.1 Algorithm with two best α-fair users

Consider an iterative algorithm with two main steps: *(i)* choose two users, i_1 and i_2 with the highest α-fair utility given by equations (5.27-5.28); *(ii)* Select the best user or both users according to the forecast scheduling rule for $n = 2$ for the time interval $[|1, T|]$ given by equations (5.20), (5.21) and (5.26). The algorithm is depicted in Table 5.1.

$$i_1(t) = argmax_{j \in users} \frac{\phi(S_j^t)}{R_j(t)^\alpha} \tag{5.27}$$

$$i_2(t) = argmax_{j \in \{users\} - \{i_1(t)\}} \frac{\phi(S_j^t)}{R_j(t)^\alpha}, \tag{5.28}$$

where $R_j(t)$ is the mean past data rate (calculated over some time window) of the user j until time t.

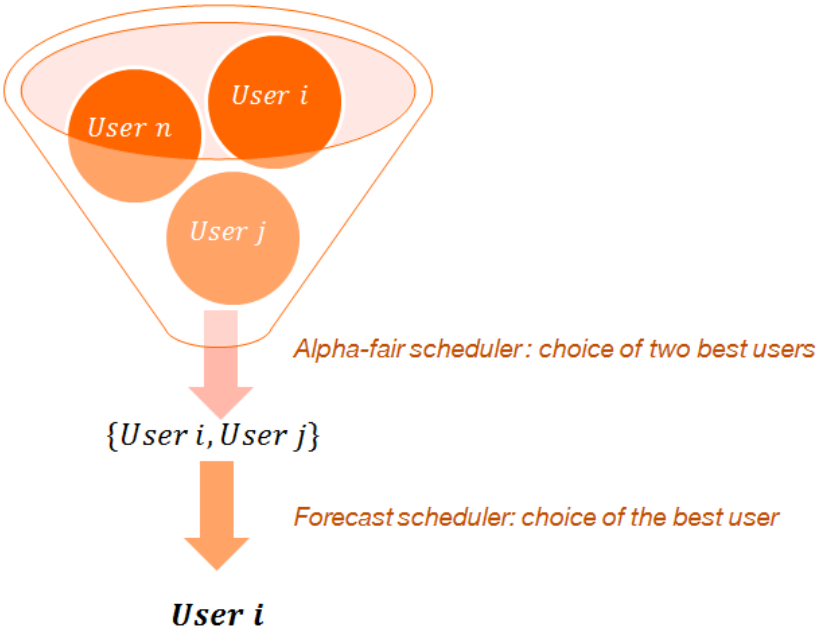


Figure 5.4: Scheduling using the best two $\alpha-$fair heuristic

The user selection given in Table 5.1 is sub-optimal with respect to the direct Forecast Scheduling technique (5.3) since it considers only the future SINR trajectories of the two best $\alpha-$fair users at each iteration. Interestingly, this approach provides significant gains with respect to RR scheduler.

Table 5.1: TBUA

Algorithm of approximation with two best α-fair users	
Step 1	-Select the two best users i_1 and i_2 according to (5.27) and (5.28) 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 (5.23). If it exists therefore use strategy (5.20)-(5.22). Otherwise use (5.26), and select the user to schedule at time t.
Step 4	-Refresh mean past data rate for all users by taking into account the data rate received at time t
Step 5	-Proceed to time $t+1$ while t verifies the following condition comparing to the scheduling duration T: $T \geq t+1$.

5.6.2 Algorithm with two clusters

The second algorithm (see Table 5.2) consists in splitting all users into two clusters. An example is a cluster for fixed users and a second one for mobile users, as shown in Fig.5.5. Clustering can be done according to other criteria such as the type of trajectories or services.

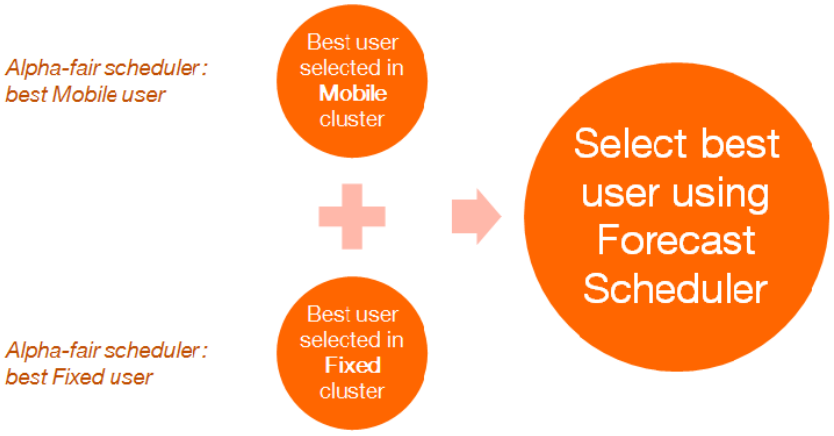


Figure 5.5: Selection methodology of best user clusters

The best user of each cluster is first selected using a standard $\alpha-$fair sched-

uler. Then a single best user is selected using the forecast scheduler. Denote by i_M the best mobile user in the mobile users cluster, and i_F the best one in the fixed users cluster:

$$i_M(t) = argmax_{j\in\{Mobile\ users\}} \frac{\phi(S_j^t)}{R_j(t)^\alpha} \tag{5.29}$$

$$i_F(t) = argmax_{j\in\{Fixed\ users\}} \frac{\phi(S_j^t)}{R_j(t)^\alpha}. \tag{5.30}$$

Table 5.2: Two best users of the clusters algorithm

Algorithm of approximation with two users'clusters	
Step 1	-Regroup users into two clusters: mobile cluster and fixed cluster at time t;
Step 2	-Select the best user in each cluster (i_M for mobile users and i_F for the fixed ones) thanks to (5.30) and (5.29);
Step 3	-Apply KKT resolution for the case of two users for i_M and i_F. Verify if time K exists by using (5.23). If it exists use strategy (5.20)-(5.22). Otherwise use (5.26), and select the user to schedule at time t.
Step 4	-Refresh mean past data rate for all users by taking into account the data rate received at time t
Step 5	-Proceed to time $t+1$ while t verifies the following condition comparing to the scheduling duration T: $T \geq t+1$.

The above two algorithms, in addition of being fast, implicitly take into account users arrival and departure, random mobility and possible update of SINR map. The problem of SINR prediction error may distort the optimality of the users' selection by the Forecast Scheduler as discussed presently.

5.7 SINR prediction error

The forecast scheduler utilizes SINR values provided by a REM. These values are averaged and depend on the interference level during the measurement time, which in turn depend on the loads of the interfering cells. We expect that the dominant contribution to the difference between the actual SINR experienced by the user and the predicted value is interference. This difference is denoted as the prediction error. Example of other sources of errors include fast fading, and interpolated errors of the REM. We consider hereafter only the error due to interference.

In order to evaluate the impact of the prediction error on the performance of the forecast scheduler, we assume an extreme (pessimistic) case: The REM provides SINR predictions for an overloaded network, i.e. the interfering cells have maximum load that equals 1. The forecast scheduler performance will be compared with that of a network with low mean load value, and is denoted as *actual network*. The interference in the actual network varies in time and the dynamic interference fluctuations is simulated at each scheduling iteration.

Interference modeling: We assume elastic traffic model in which all the BS resources are allocated, even if there is just one user. At any time, a neighboring cell can have one of two states: it can be empty and produce no interference; or it has at least one user and produce maximum interference. The interference term is therefore described as follows:

$$I = \sum_{i=1}^{n} 1_{(N_i>0)} \frac{P_i h_i}{r_i^{\gamma}}, \tag{5.31}$$

where N_i is the number of users in the neighboring cell i, P_i - the transmitted power, h_i - the channel gain including the path loss and antenna gain, and r_i - the distance from the base station i to the desired location. We rewrite $1_{N_i>0} = V_i$ where V_i is a Bernoulli random variable with parameter p:

$$I = \sum_{i=1}^{n} V_i \frac{P_i}{r_i^{\gamma}}. \tag{5.32}$$

All the V_i have the same distribution, but typically they are not independent due to interference among cells. However we suppose that V_i are independent due to the small impact on the interference estimation. p is defined as the mean load of the cell.

By comparing the throughput gain achieved using actual SINR values, one can estimate the impact of the prediction errors using the following formula:

$$error = \frac{1}{n \times T} \sum_{i \in users} \sum_{t \in times} 1_{\{|a_i(t) - ae_i(t)| > \epsilon\}}, \tag{5.33}$$

where a is the forecast scheduling strategy using the REM, namely the predicted SINR values and ae is the forecast scheduling strategy using measurements from the actual network and ϵ - the tolerated error.

5.8 Users impatience problem

Using this FS takes advantage of future high data rates or preserves from low ones. However a user may wait some time before he is served due to the scheduling duration T. Some users have no patience to wait to be served and could leave the network which would deteriorate the user experience.

We may reduce the scheduling delay for the forecast scheduling. For instance we could take $1ms$ instead of 2 ms in scheduling delay (step). This solution is improper because it does not change the forecast scheduling strategy specially when the mean value of the SINR follows a continuous trend. If there is no sufficient variation of signal power (continue function) between each scheduling delay, there is no need to reduce this delay from a certain minimum. However, if there is no SINR trend, this solution may be interesting. In general, there is always a signal tendency along the user trajectory. A solution to the impatience problem is to reduce the FS duration T. This duration should depend on users' impatience time which is random variable.

Based on empirical studies carried out by Selvidge ([53] and [54]), a threshold of 30 seconds was suggested for elastic traffic. Other thresholds are proposed in [55], [56]. Different reasons may vary the waiting time of a user (patience time): the speed of the internet connection, the local network traffic, the data volume requested to download, the type of the requested data to download etc.

Depending on their data requested volume, we separate users in different clusters. Each cluster has a typical waiting time computed as follows: a user remains patient with probability of 99% if the random variable γ of user's time patience is below an unknown threshold β:

$$\mathbb{P}(\gamma < \beta) = 99\%, \tag{5.34}$$

where β is the maximum patience duration that a user stays in the network with probability 99%. A similar approach is given by [57].

The forecast scheduling duration T should therefore be fixed to β. β can be approximated empirically based on historical data of users impatience duration [58]. By classifying historical patience durations of users (thousands of data items) in increasing order in a table, β is then taken as the patience time located at 99% line from the bottom of the table (see Fig.5.6). Other β computation techniques (more accurate and applied in finance field to evaluate the maximal loss (Value-at-Risk VaR)[59], [60], [61]) may be considered if we know the distribution of the random variable γ or we do not have sufficient number of historical data items.

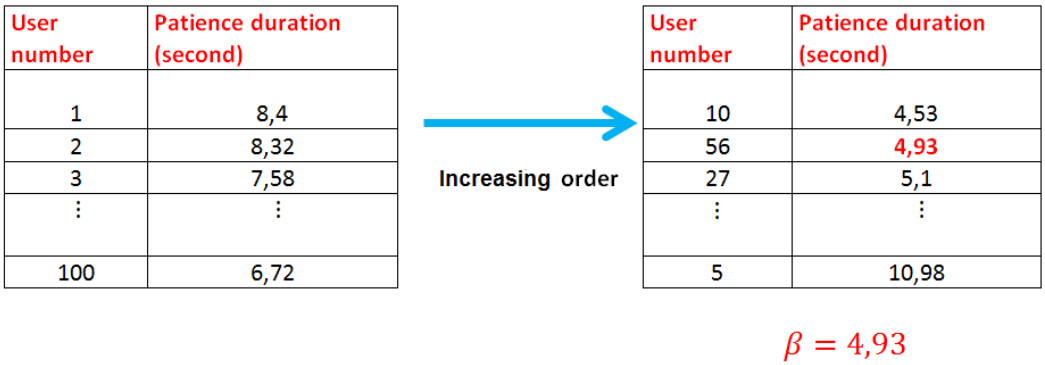


Figure 5.6: Selection of the maximum patience duration γ duration which a user stays in the network with probability 99%

After duration T, we reiterate the forecast scheduler for the time window $T = \beta$. However taking a small T may some times not give an optimal forecast scheduling strategy (see numerical section). We can for example decrease the probability of leaving the network from 99.5% to 98%.

5.9 The problem of random events

This Section generalizes the optimization model (5.3) in order to take into account arrival and depatures of users as well as different user classes such as fixed and mobile users, real time and non-real time services, etc.

5.9.1 Multiclass problem

At any time, any user may arrive with a particular request from the network: downloading or uploading data, sending SMS, calling or watching video Youtube, etc. One cannot treat all the users in the same manner and should

cluster them in terms of the service type requested. We choose to regroup users by the type of services: real-time (call, Skype, some video streaming), and non-real time users (download and upload data, HTTP, etc.).

Denote by C_1 the set of real-time active users and C_2- the set of non-real time (elastic) users. Let T_1 be the minimum scheduling period for a real time user. T_2 is the forecast scheduling duration for all users in class C_2, which is subdivided into minimum scheduling periods of T_1 (namely $T_2 >> T_1$).

Static or low speed users can be included in the set C_1 where users can benefit from fast fading. It is noted that if service differentiation is sought, weighting coefficient $w_{s(i)}$ ($s(i)$ is the service of the user i) can be introduced as in Weighted Fair Queueing scheduling. The new forecast scheduling model including the real-time set is as follows:

$$\begin{aligned} maximize\ f(a) = & \sum_{i \in C_2} \frac{(\sum_{t=1}^{m} a_i(t)\phi(S_i^t))^{1-\alpha}}{1-\alpha} + \\ & \sum_{i \in C_1} \sum_{t=1}^{m} w_{s(i)} \frac{(a_i(t)\phi(S_i^t))^{1-\alpha}}{1-\alpha} \\ & with : \forall i, \forall t, a_i(t) \geq 0 \\ & and : \forall t, \sum_{i \in C_1 \cup C_2} a_i(t) = 1 \end{aligned} \tag{5.35}$$

where $m = T_2/T_1$.

We can have as much classes as type of users and type of trajectories. For example users approaching to a traffic junction with a distance below d are attributed to a new class, denoted as C_3 in Fig.5.7. The corresponding duration T_3 for users in C_3, $T_3 < T_2$, is defined as a function of d. The function that we add to the optimization problem (5.35) is a Restricted Forecast Scheduling (RFS) during time T_3. This time is called a stopping time when the users draw near at the random trajectory point (see Fig.5.7) a crossroad for example. These users as soon as they are next to the random point with a distance less than d are switched over to the set C_1 due to both lower speed and trajectory uncertainty.

In the following table (Table 5.20) we summarize the considered type of schedulers with respect to users type of mobility and channel behaviours.

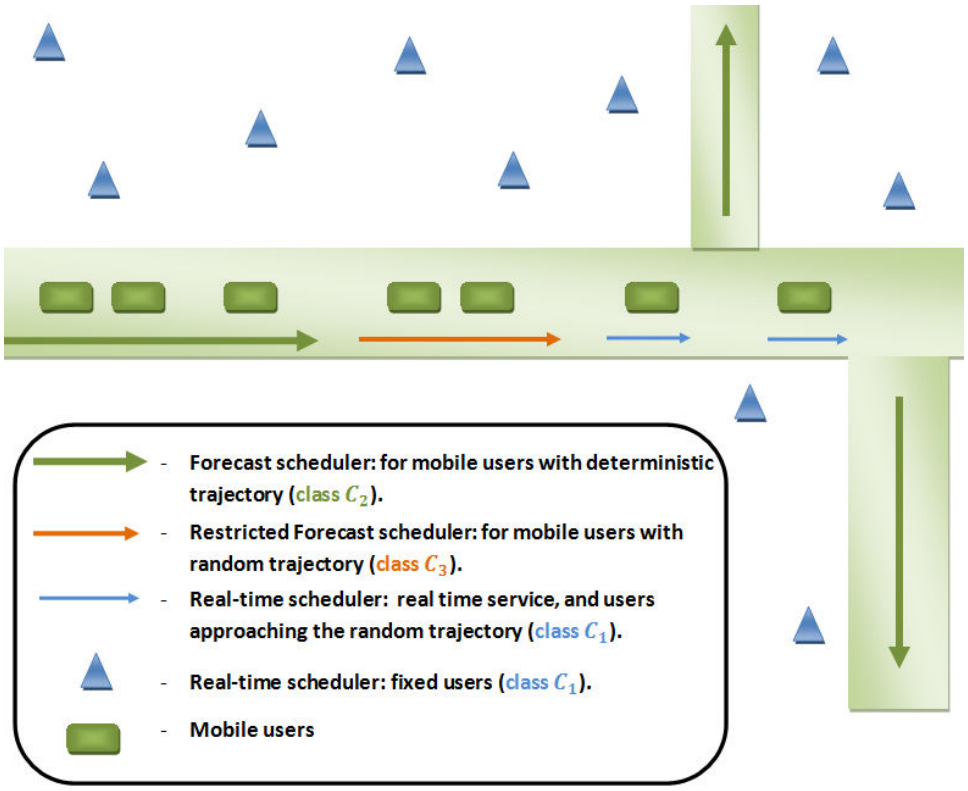


Figure 5.7: Forecast scheduling with different traffic classes

Table 5.3: Scheduling types with respect to users' mobility

User type	**mobility behavior**	**channel behavior**	**Scheduler**
High speed	Deterministic speed and trajectories	short coherence time	Forecast Scheduling C_2
Middle speed	Random speed	short coherence time	Restricted Forecast Scheduling C_3
Low speed	Random trajectories	fading gain	simple $\alpha-$fair scheduler C_1
Fixed	Deterministic	fading gain	simple $\alpha-$fair scheduler C_1

5.9.2 Arrival and departure problem

In the real network, users may arrive and leave the network at any time. The FS of the optimization problem (5.3) does not consider the random event oc-

currence such as arrivals, departures, SINR fluctuations, trajectories variation, etc. For example a user that arrives during the FS period T is not considered in the resource sharing. This is not fair for him and he may wait some seconds until he is considered by the network if he was patient.

The extension of the optimization model (5.3) to include arrivals and departures is described presently. One can make the same extension for the optimization problem (5.35). We could integrate new users in the set C_1 at each arrival time, however they will not benefit from the forecast scheduling (i.e. if users were in high speed).

The solution consists of initializing the optimization problem (5.3) or (5.35) at each arrival and departure. To remain fair among all users, including those who have not yet been scheduled, we recompute the forecast scheduler at each event but taking into account the past data received. This new heuristic scheduling is called the Updated Forecast Scheduling (UFS) and is inspired by the rolling horizon approach [62]. It is noted that optimizing the problem (5.3) or (5.35) for each time step or for each random event occurrence time (random variation of speeds, trajectories, number of users, SINR map) gives the same result, however the latter has significant lower compelexity.

This equivalence is due to the fact that if no events occur then the new initialized optimization problem, with recall of past received data, gives the same results at each time step during T as between time events. It is important to reinject at each time the data received in order to stay fair for all the remained users. In Fig.5.8 we explicit the scheduling algorithm for the particular optimization problem (5.3):

1. At the initial time $t = 1$ we maximize the function f of the optimization problem (5.3) denoted by f_1;

2. We find $b^1 = (b_1^1(1), ..., b_1^1(T), b_2^1(1), ..., b_2^1(T), ..., b_n^1(1), ..., b_n^1(T))$ the optimal solution of the function f_1;

3. For the second event (arrival for example) denoted by time $t = u_2$, we replace $a_i(1), ..., a_i(u_2 - 1)$ by $b_i^1(1), ..., b_i^1(u_2 - 1)$ for all users i in the function f_1. We have a new function f_2 to optimize at time $t = u_2$ (see Fig.5.8).

4. We proceed iteratively the optimization for the function f_N at an event number N for $N \in [1, T]$:

$$f_N(a) = \sum_{i=1}^{n} \frac{(\sum_{t=1}^{u_N - 1} b_i^t(t)\phi(S_i^t) + \sum_{t=u_N}^{T} a_i(t)\phi(S_i^t))^{1-\alpha}}{1-\alpha} \tag{5.36}$$

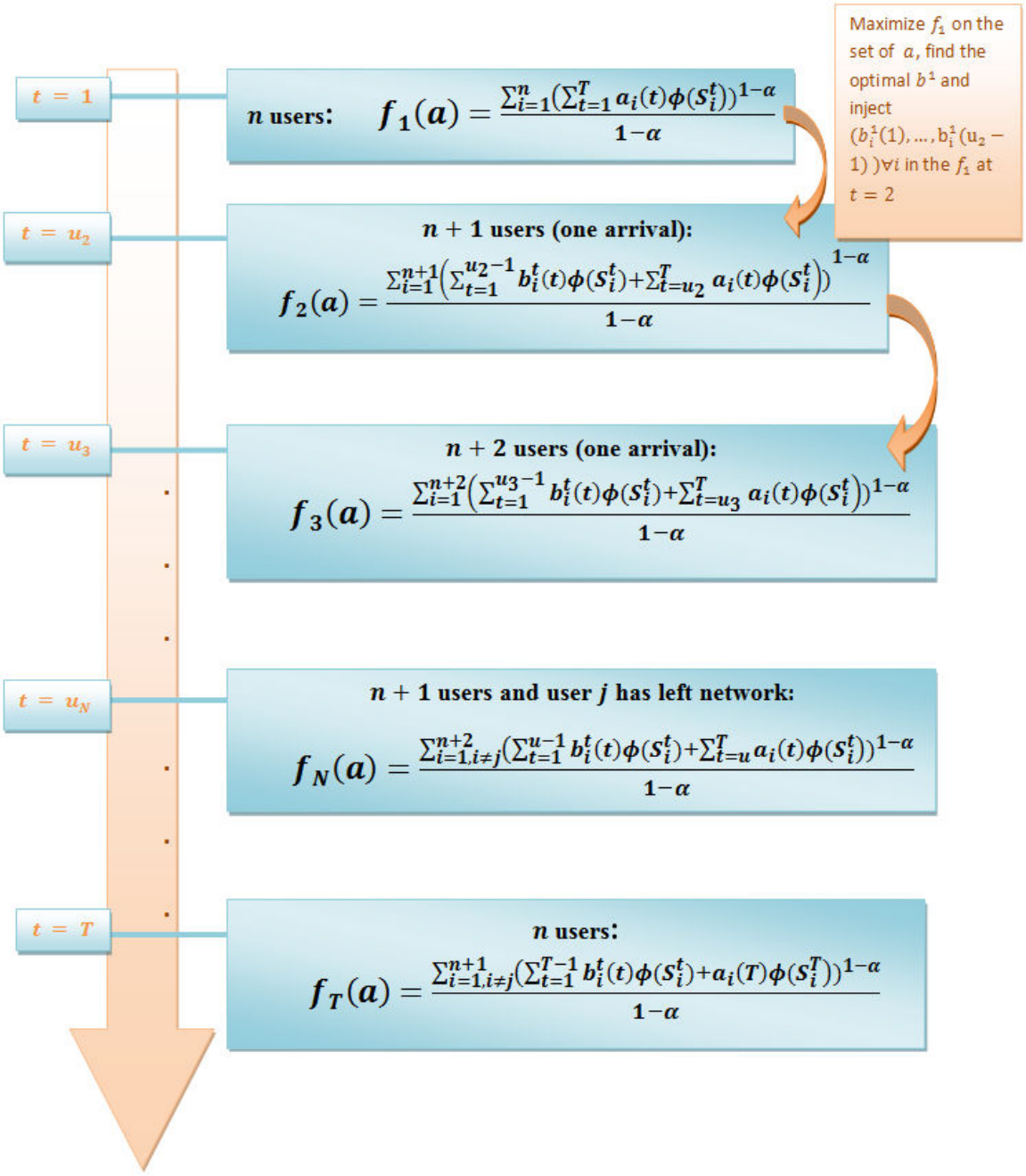


Figure 5.8: Algorithm of the forecast scheduler taking into account arrival and departure of users

5.10 Numerical results and simulation

We present in this section the numerical simulations that have been carried out using our Matlab simulator. Users are moving with a speed of 50 km/h. We recall that since at this speed, due to too short coherence time, one cannot exploit fast fading to achieve opportunistic scheduling gain, and for this reason we consider the RR as a base line to the different approaches.

5.10.1 CVX resolution

The nonlinear objective function of the optimization problem (5.3) is a convex function. In this case a modeling system for solving convex programs with constraints is needed. We propose in this thesis the **CVX** library implemented in Matlab to resolve this kind of problem (see [63] and [64]). We have particularly chosen this resolution algorithm for sake of simple usability and also the authors of this algorithm have given the algorithm as an open source for free in their internet page (http://cvxr.com/cvx/). The **CVX** resolution process verifies the convexity of the problem and solves it using **SDPT3** or **SeDuMi**. **SDPT3** is a MATLAB implementation of infeasible path-following algorithms for solving conic programming problems whose constraint cone is a product of semidefinite cones. It uses a predictor-corrector primal-dual path-following method, with different types of search direction. **SeDuMi** is a linear/quadratic/semidefinite solver for Matlab and Octave. We will discuss this solution and compare it to other algorithms in the Numerical results section.

5.10.2 Forecast scheduling and TBUA gain

We study a system with 10 vehicular users following each other with distance between them of 10 m (see section 5.4). We compare the FS model (5.3) with the Two Best Users Algorithm (TBUA) (Algorithm 1-table 5.1) and the RR scheduler with highest interference. As mentioned earlier, we suppose that the SINR is provided to the FS via a REM. To make the comparison *fair* we suppose that the SINR cartography has been produced for a *pessimistic* scenario where interfering cells are fully loaded. Hence a maximum difference between the actual SINR measurements and the REM input is considered. Interestingly, it is shown that the this difference in SINR has little impact on the forecast scheduler decisions.

The characteristics of the traffic and the network are described in the Table 5.4.

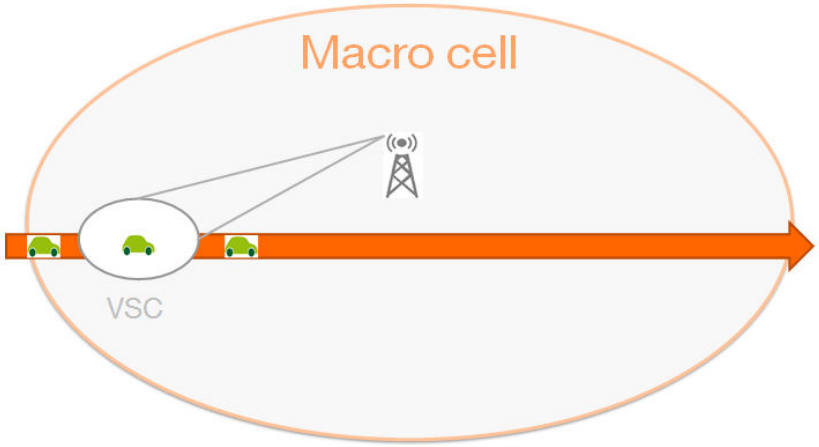
Table 5.4: Network and Traffic characteristics

Network parameters	
Number of macro BSs	1
Number of interfering BSs	6
Macro-cell layout	hexagonal omni sector
Intersite distance	$500m$
Bandwidth	$20MHz$
Channel characteristics	
Thermal noise	$-174dBm/Hz$
Macro Path loss (d in km)	128.1 + 37.6 $\log_{10}(d)$ dB
Mobility traffic characteristics	
User speed	$50km/h$
Number of users	10
Time between vehicles	$1.4s$
File size σ	full buffer (∞)
Error tolerated ϵ (eq.5.33)	0.05
Forecast scheduling duration T	$20sec$
One iteration = Scheduling delay	$70ms$
Average sliding window	$15iter.$

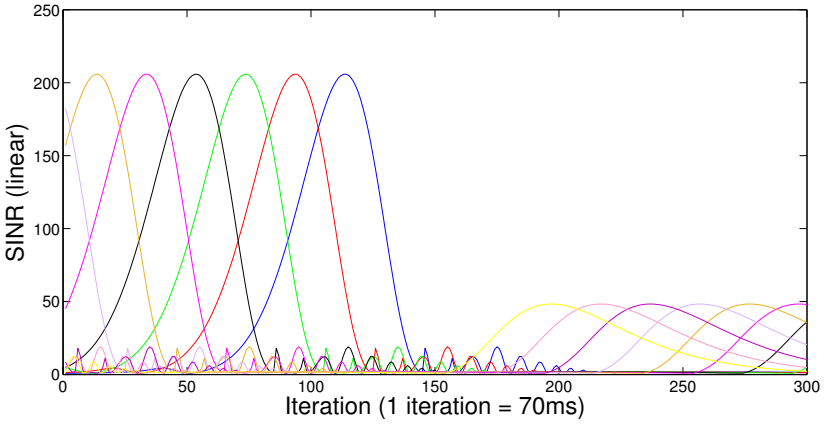
Throughout this section we suppose that there is only mobile vehicular users (Algorithm 5.2 for the two user clusters is therefore excluded). Three mobility scenarios are investigated to assess the impact of SINR diversity along the trajectory on the FS performance. In two of the scenarios, a VSC is used in order to introduce high SINR variations:

- **Scenario 1**: a VSC (fixed beam) is located beside the road close to the cell edge (Fig.5.9);
- **Scenario 2**: vehicular users drive along the road crossing the cell far away from the VSC. (Fig.5.10);
- **Scenario 3**: two users are driving along two parallel roads with - and without a VSC (Fig.5.11).

In this Section the scheduling gain is calculated with respect to RR scheduler with neighboring cell load of 10%. Fig.5.12 compares the scheduling gain in terms of MUT with respect to RR scheduling for the FS and the TBUA. The gain for both FS and the TBUA is significant for all scenarios. The FS

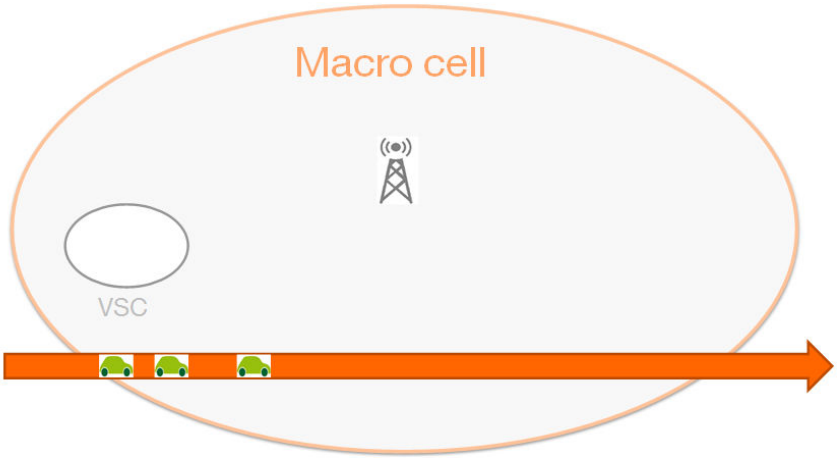


(a) Trajectory with VSC deployment

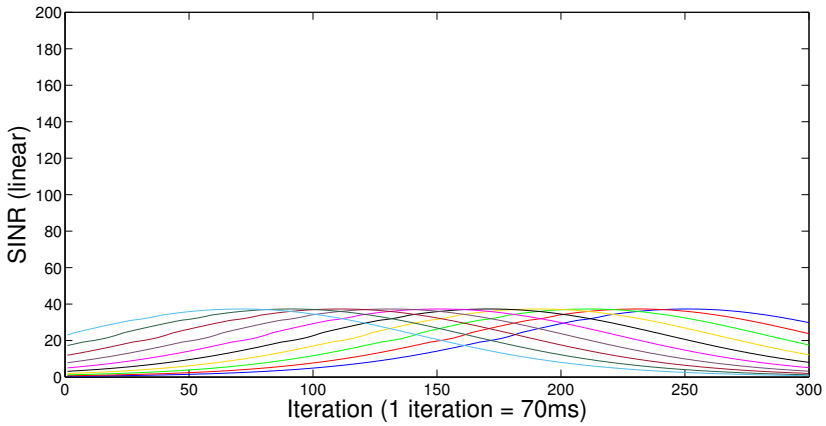


(b) SINR for different users along the road

Figure 5.9: Scenario 1 trajectories

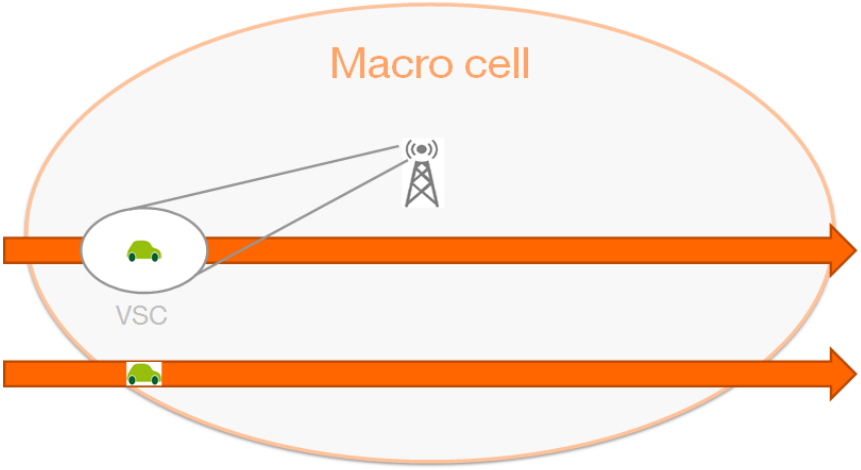


(a) Trajectory without VSC deployment

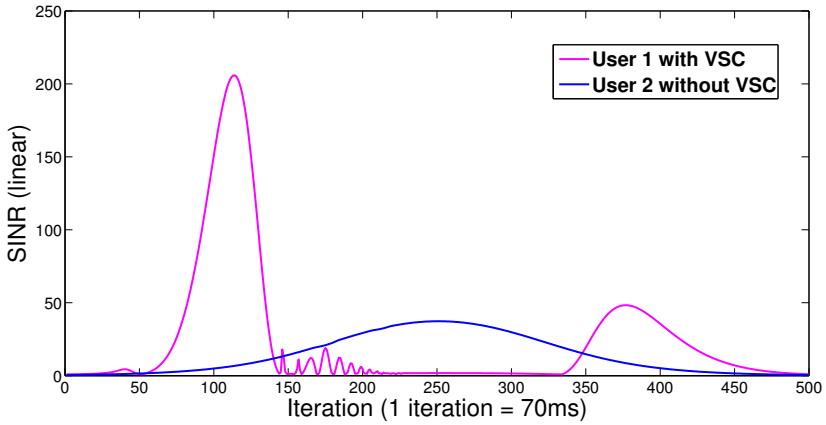


(b) SINR for different users along the road

Figure 5.10: Scenario 2 trajectories



(a) 2 users trajectories



(b) SINR for different users along the road

Figure 5.11: Scenario 3 trajectories

gain is of 117%, and for the TBUA of 80%. Scenario 2 achieves the lowest scheduling gain of 30% and of 15% for the FS and the TBUA respectively. Inspite of the low SINR variations along the trajectory, significant gains are still obtained.

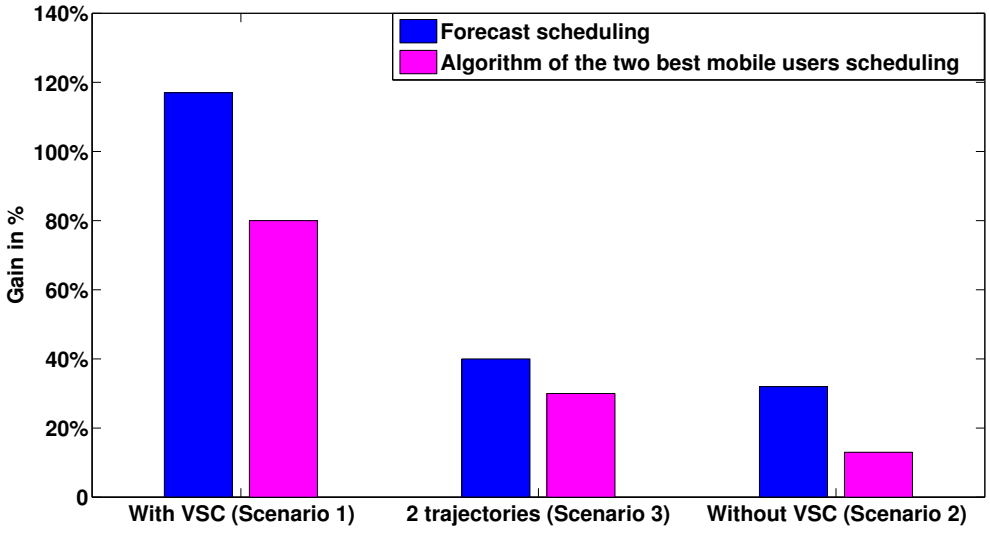


Figure 5.12: Forecast scheduling and two best α-fair user algorithm gain comparing to RR scheduling for the three scenarios.

5.10.3 Impact of interference error

We investigate the interference impact on the FS for the scenario 3 (see Fig.5.11). We recall that the REM data corresponds to maximal interference (neighboring cell load of 100%). A maximal interference means that all the cells have at least one active user. We compute the difference in scheduling decisions (error) using eq. (5.33)) for scenario 3 in the case where the neighbor cells' load is of 10%. In Fig.5.13, one can notice that if interference changes randomly in each iteration with average neighbor cells' load of 10%, the difference in scheduling decisions reaches 22% (cyan curve).

The difference in scheduling decisions decreases with the increase of the duration T of the FS. When the scheduling interval is too short, the SINR dynamics is likely to be limited, and the diversity in the mean SINR along the trajectory can be less exploited. In this case, the impact of interference variations on the scheduler strategy (decisions) will be higher.

Fig.5.14 shows that the gain of the FS in mean data rate comparing to RR increases as a function of the scheduling duration T. One can see that the difference scheduling decisions in the FS (5.3) due to significant change in interference translates to negligible difference of the data rates. This is a key point that characterizes the FS solution that is illustrated below. The more we have information in time about the SINR trajectories of users, the more we gain in predictive scheduling. Figs. 5.15 and 5.16 describe how the mean user data rate varies in time for the two extreme cases: $T = 2s$ and $T = 20s$. A 15 iterations sliding window (i.e. of arround one second) for the plots is considered. Four scheduling strategies are compared: FS (5.3) with interference error, FS without errors, approximation with the TBUA, and RR scheduling. One can see that the interference error on the MUT achieved by the FS is negligeable for both scenarios of $T = 2s$ and $T = 20s$. It is interesting to see that the less is the duration T, the more TBUA approaches the FS (5.3).

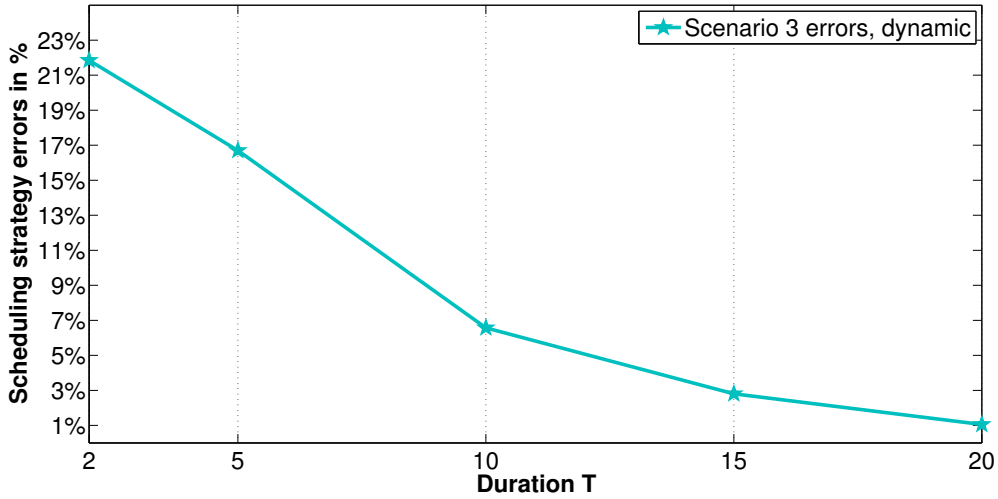


Figure 5.13: Scheduling decision difference (error) for 10% interfering cell load (eq.(5.33)) for scenario 3

5.10.4 Impact of random events and trajectories' uncertainty

Two different types of randomness are studied in the following: trajectory randomness and a user arrival in the network. In this section we analyze the impact of a random event occurrence on five type of schedulings:

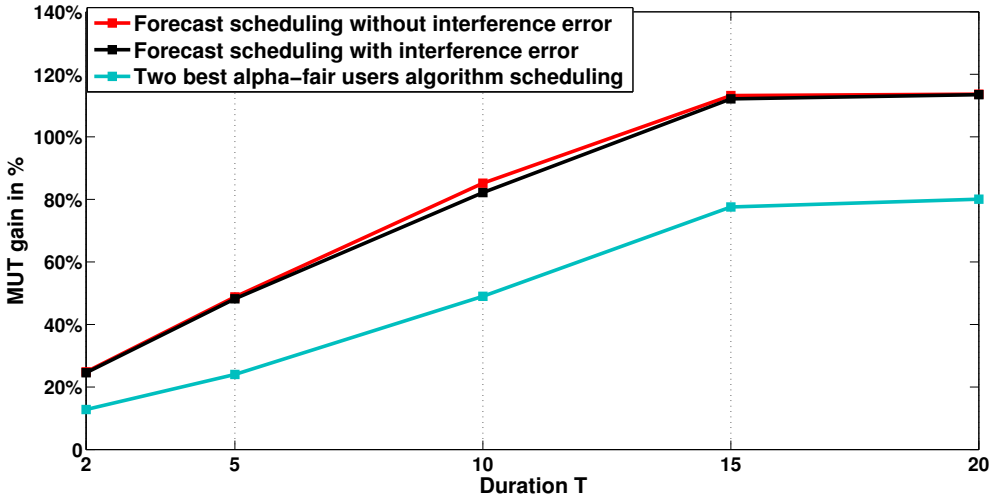


Figure 5.14: Forecast scheduling MUT gain comparing to RR with and without interference error with 10% load for interfering cells

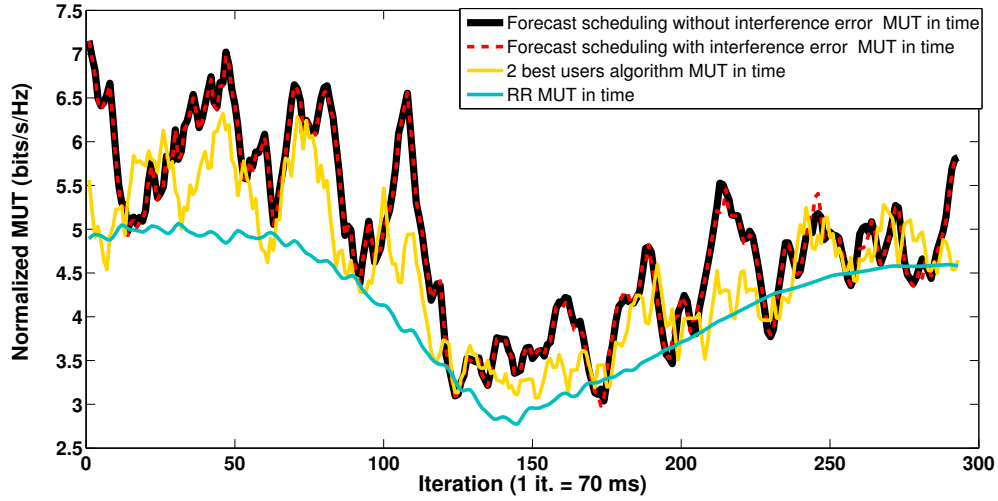


Figure 5.15: Users' throughput variation during time in case of Forecast scheduling duration T=2s, averaged with 15 iterations sliding window

- The basic FS which does not take into account any random event during the scheduling duration T (section 5.5.1);
- UFS which stops at each random event occurrence and integrates the

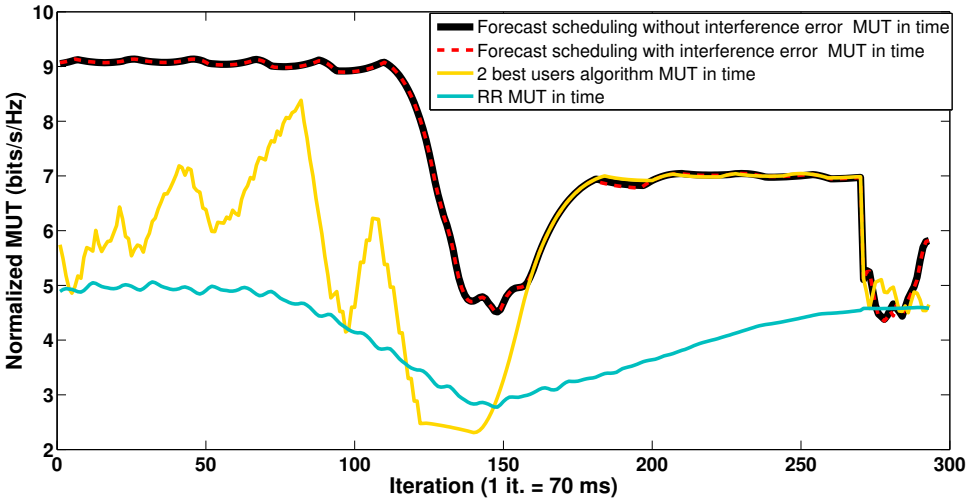


Figure 5.16: Users' throughput variation during time in case of Forecast scheduling duration T=20s, averaged with 15 iterations sliding window

past data allocated by the FS before reiterating (section 5.9.2);

- RFS which stops at $T' < T$ and integrates the past data allocated by the FS before reiterating. This scheduling is usefull when we know at which time the random event will occur (any randomness along the trajectories such as a crossroad) (section 5.9.1);
- Seer Forecast Scheduling (SFS) (is a kind of oracle) that sees the future and knows when and what will occur in the future. This last scheduling is used to compare how *perfect* are the others cited schedulings and has same formulation as the basic FS knowing all the random events.
- RR scheduling.

The two schedulers FS and UFS assume the following hypothesis on all users' trajectories: *All the users will take the Road 1 with the VSC* (Fig.5.17).

Trajectories randomness

Each vehicular user may change the road given by the GPS if for example his GPS has not seen a physical congestion in some area (due to an accident). The FS which integrate in the scheduling computation the trajectory information

given by the GPS will be in this case incorrect and not optimal. We compare here the five schedulings previously presented for this case.

Deterioration of SINR

We take for example of trajectory randomness the Fig.5.17. The figure presents users arriving to a cross road. If the users continue straight, they reach the coverage zone of a virtual small cell (*Road 1* in the Figure), whereas if they turn right, due to propagation condition they experience signal attenuation of the order of $10dB$ (*Road 2* in the Figure).

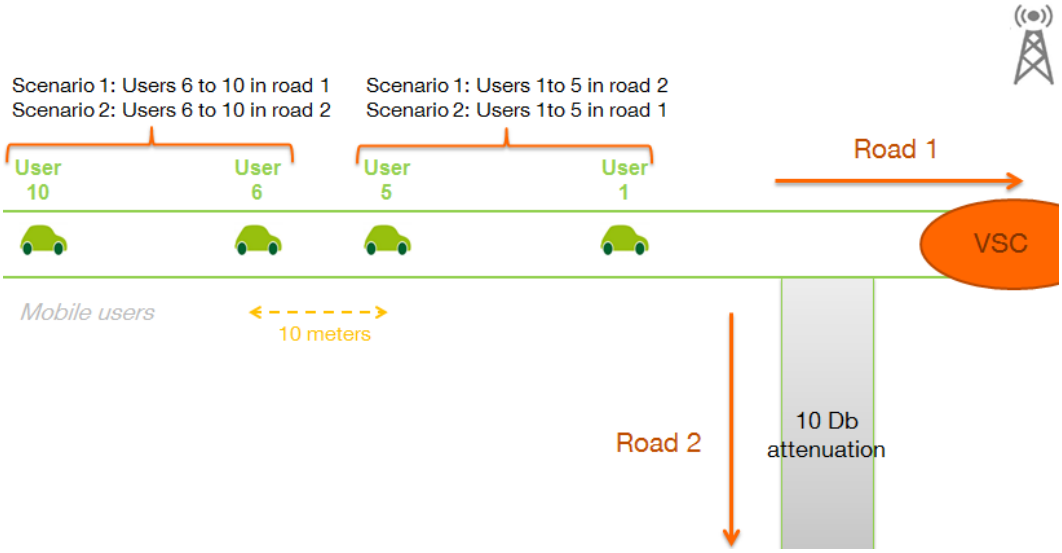


Figure 5.17: Scenarios 1 and 2 of random trajectories

Figs.5.18-5.20 depict the data received without bandwidth factor W for each user during T for the five scheduling strategies.

The basic FS (yellow bar in Figs.5.18-5.20) and the UFS (black bar in Figs.5.18-5.20) do not take into account this randomness and see only the virtual small cell. The UFS takes into account the hypothetical users trajectory (Road 1 in this example). This scheduler will stop as soon as a user takes a trajectory different than the one assumed by the scheduler. It will then takes into account the signal deterioration of the Road 2. The SFS has already integrated the future SINR deterioration for the five first/last users depending on the scenario considered.

In Fig.5.18, the five first users in the trajectory, who arrive first to the random point, turn to the road 2 (we do not know this information) and the

last 5 users take the road 1 with VSC. We denote this case by scenario 1. The five first users will not receive as much as the five last users. Even the SFS (red bar) cannot be as much *fair* for the users, as the five first users do not have much time to be scheduled before taking the road 2 (bars "before crossroad" in Fig.5.18). The FS give therefore all the resource to the five first users before the reaching the crossroad as the scheduler knows that there is a high deterioration after the crossroad in the road 2.

The five last users do not receive anything before in case of FS, UFS and SFS, the schedulers will wait until they arrive to the road 1 of the VSC. In the figure "After crossroad" of Fig.5.18, the four forecast schedulings give pratically all the resource to the five last users as these schedulers have supposed in the beginning that these users will take the VSC road. The UFS gives the best *fair* amount of data for each user. The sum of data received by each user for scenario 1 is depicted in Fig.5.19.

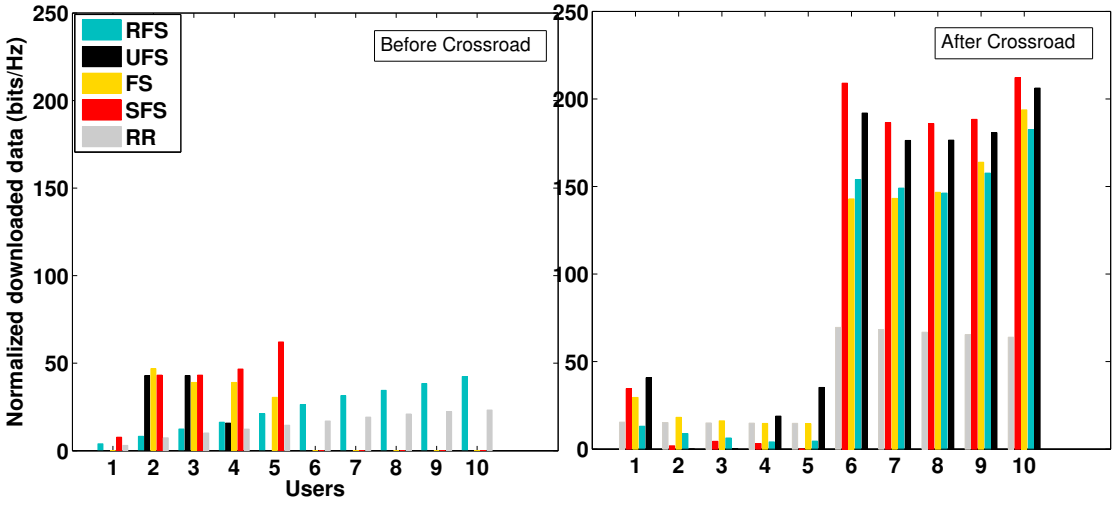


Figure 5.18: Impact of random trajectories with different scheduling strategies for scenario 1, before and after the crossroad for each user

In Fig.5.20, the five last users in the trajectory, who arrive last to the random point, are taking the road 2 (we initially do not know this information). We denote this case by scenario 2. The UFS and the basic FS do not give to the five last users any data (bars "Before Crossroad" in Fig.5.20). This is due to the fact that the scheduler assumes that the users will drive straight along the main road which passes by the VSC. When the five first users arrive to the VSC, the schedulers give all the resource to these users (bars "After Crossroad

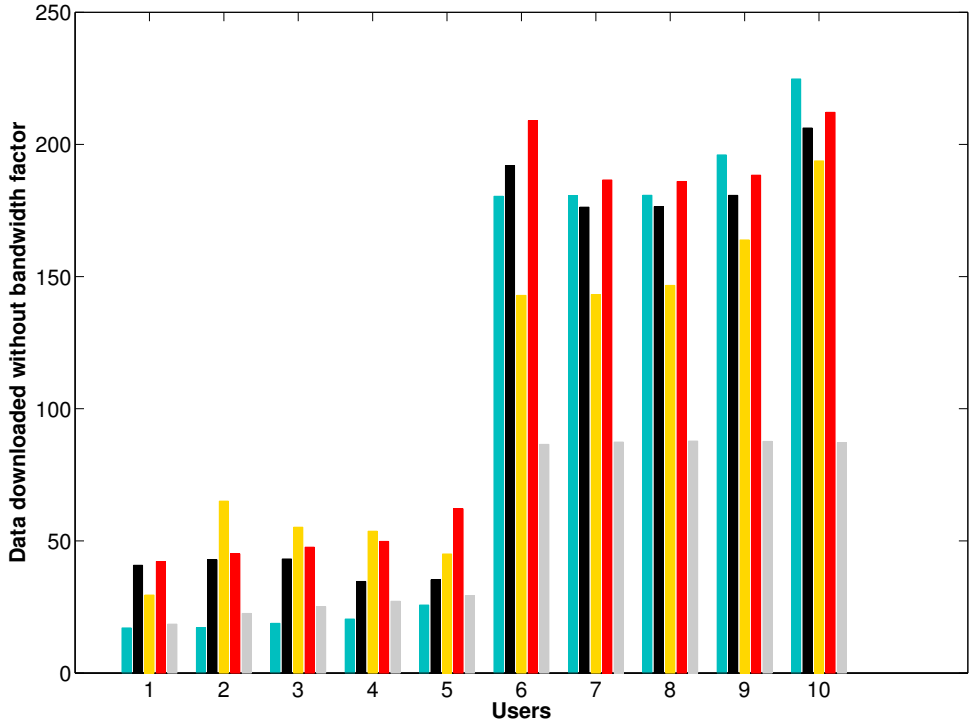


Figure 5.19: Impact of random trajectories with different scheduling strategies for scenario 1

in Fig.5.20).

The sum of data received by each user for scenario 2 is depicted in Fig.5.21. The UFS and FS have approximatly the same behavior in this scenario. The reason is that these schedulers compute the allocation resource considering a hypothesis (all the users will take the road 1).

The UFS (black bars) is therefore the best scheduler to choose in this case of trajectory uncertainty providing the best utility in terms of data rate and fairness (recall that the SFS is just an oracle theoratical scheduling).

The RR scheduler (grey bar) underperforms all the forecast schedulers but still gives a quite little data to users before they take the road 2.

If there were a tunnel instead in the road 2, the RFS would be the best scheduler to consider as the UFS will not have time to reschedule fairly after taking the tunnel. Some users will not receive anything in the case of scenario 2 considering the UFS (see the next part).

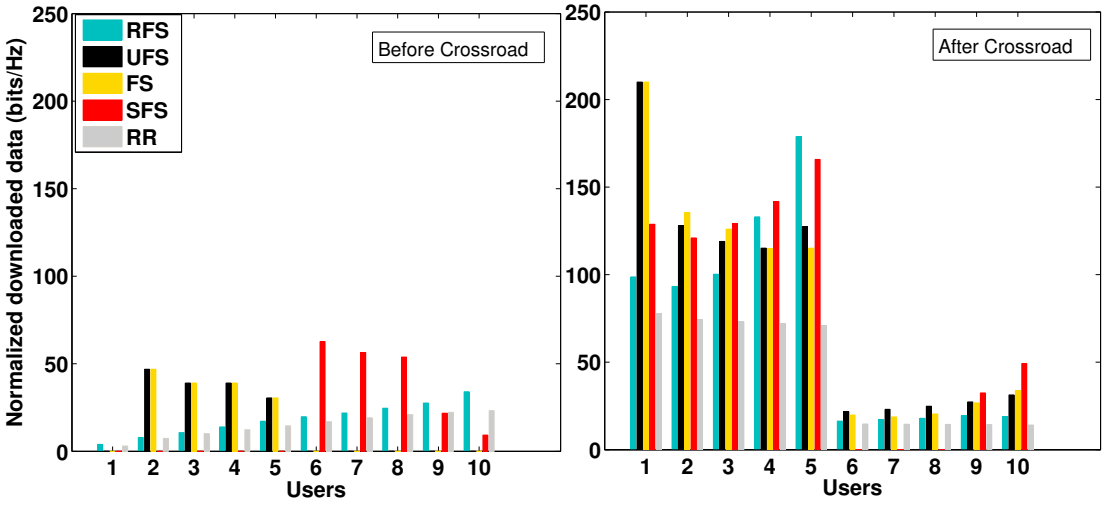


Figure 5.20: Impact of random trajectories with different scheduling strategies: scenario 2, before and after the crossroad for each user

Tunnel case

We consider the same scenarios as previously in the case of SINR attenuation (see Fig.5.17), but this time in the Road (2) there is a tunnel instead where users are disconnected. Figs.5.22-5.23 depict the data received without bandwidth factor W for each user during T for the five scheduling strategies.

In Fig.5.22, the five first users in the trajectory, who arrive first to the random point, turn into the tunnel road and the last 5 users take the Road 1 with VSC. We denote this case by scenario 1 in this part. These users will not receive as much data as the five last users. Even the SFS (red bars) cannot be as much fair for the users since the five first users do not have time to be scheduled before the tunnel. User 1 is not even scheduled with the basic FS nor with the UFS. Users 2 to 5 receive some data as they approach the VSC.

In Fig.5.23, the five last users in the trajectory, who arrive last to the random point, are taking the tunnel road. We denote this case by scenario 2 in this part. The UFS and the basic FS one do not give to the five last users any data. As explained before, the FS and UFS consider that the users continue straight along the main road. When the five first users arrive to the VSC, the schedulers give all the resource to these first users, and at the same time the last five users arrived to the tunnel or approach it and do not allocated anything.

The RFS (green bars) is therefore the best scheduler which provides the

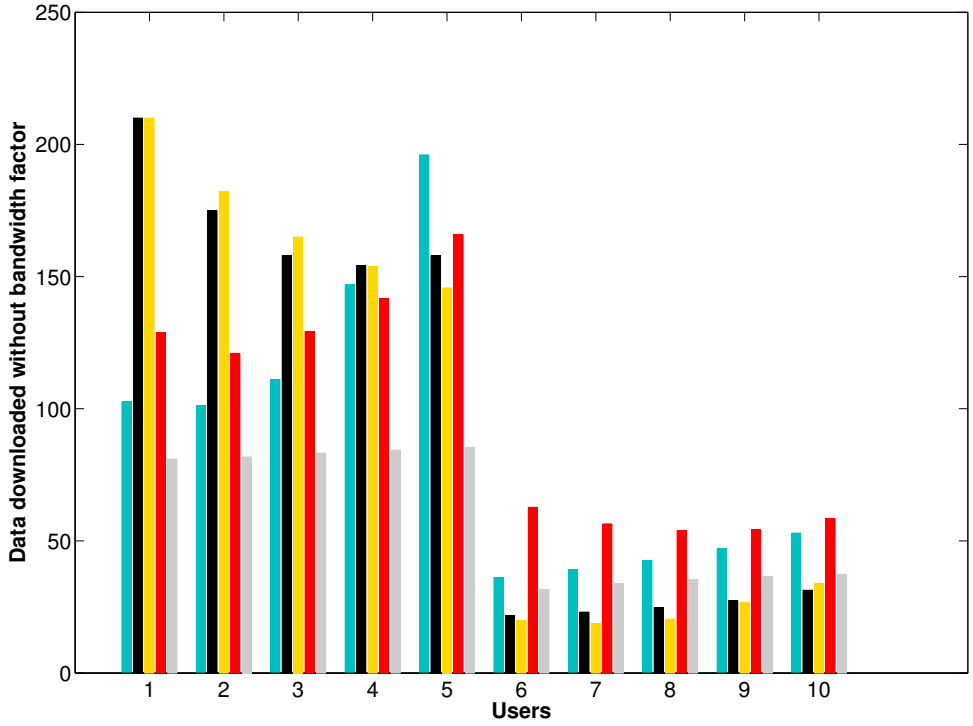


Figure 5.21: Impact of random trajectories with different scheduling strategies for scenario 2

best utility in terms of data rate and fairness.

Arrival randomness

A vehicular user may arrive in the network at any time by leaving a parking or a bus/tramway station etc. An arrival of a user at three different times, during the FS interval time T is investigated in this part. The system considers the same users model with 9 users in this case, and the arrival of user 10. A VSC is deployed in the users' trajectory with no random trajectories for all users (see Fig.5.24).

Only four scheduling strategies are considered as we do not know the event occurrence time, namely the RFS is not investigated in this part. The impact of one user's arrival (user 10) is depicted in Fig.5.25-5.27. The color code of

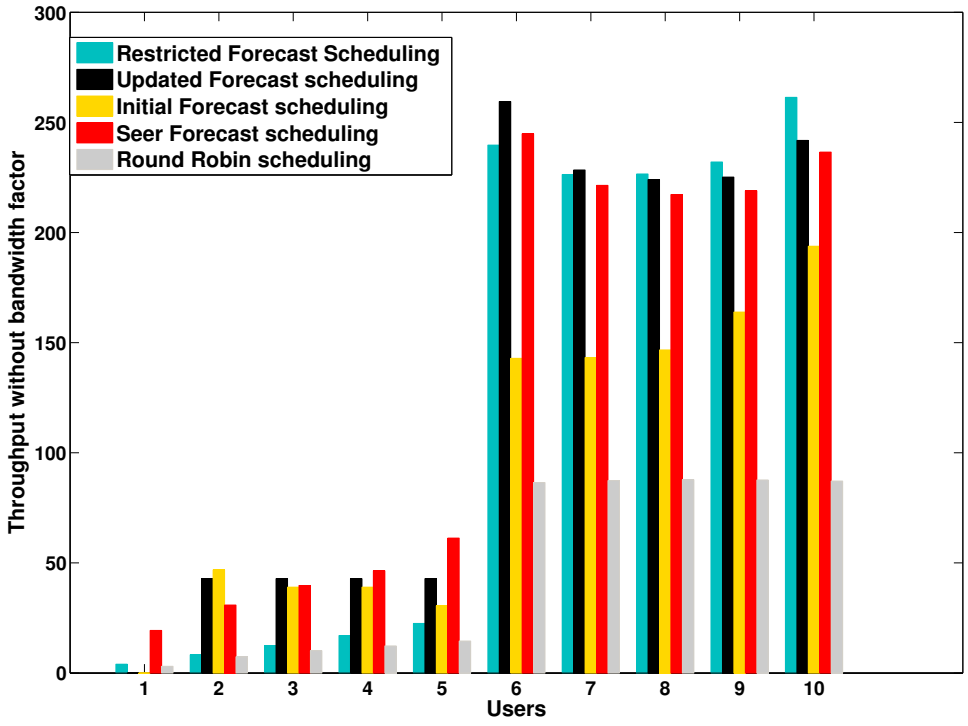


Figure 5.22: Impact of random trajectories with different scheduling strategies for scenario 1 in case of a tunnel in Road 2

the bars and the key performance studied are the same as in the section of random trajectories 5.10.4.

The difference between these figures is that the user may arrive at three different times whithin the time interval T: in the first times (Fig.5.25), in the middle one (Fig.5.26) or at the end of the forecast scheduling time interval T (Fig.5.27). Suppose here $T = 400\ it$ namely $T = 28\ s$

In Fig.5.25, the user 10 arrives at time $t = 10\ it$. The main observation is that the UFS and the SFS transmits almost the same dowloaded data during the trajectories for all the users even for the user that arrives randomly. In fact, the UFS considers the arrival as a user that has never receive any data and tries to stay fair between the new arrival and the older users present in the cell. Since the user arrival occurs in the beginning of the scheduling period, in the remaining time the UFS enforces fair and optimal allocation between the users.

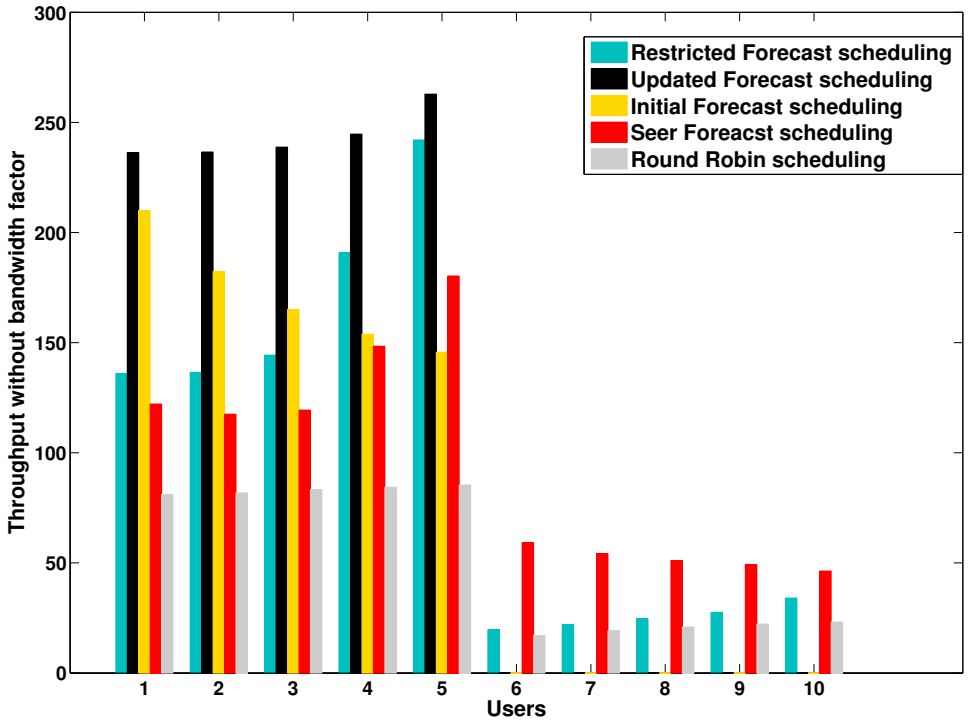


Figure 5.23: Impact of random trajectories with different scheduling strategies for scenario 2 in case of a tunnel in Road 2

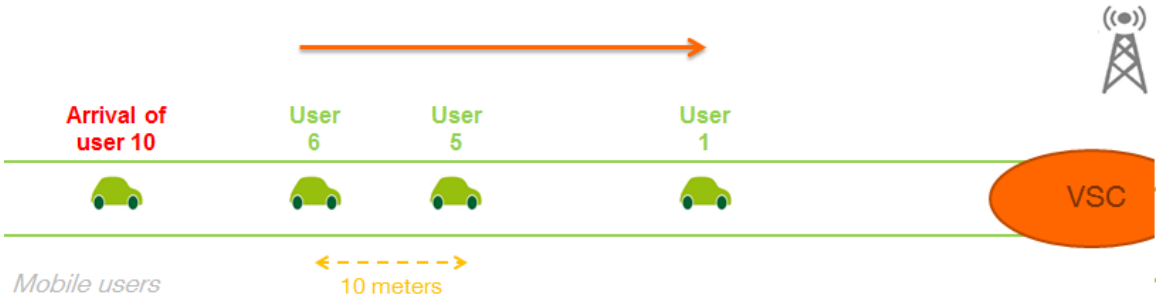


Figure 5.24: Random arrival in time of user 10

In Fig.5.26, the user 10 arrives at time $t = 200\ it$. The UFS and the SFS are different from the first case of Fig.5.25 for user 10. The schedulers do not

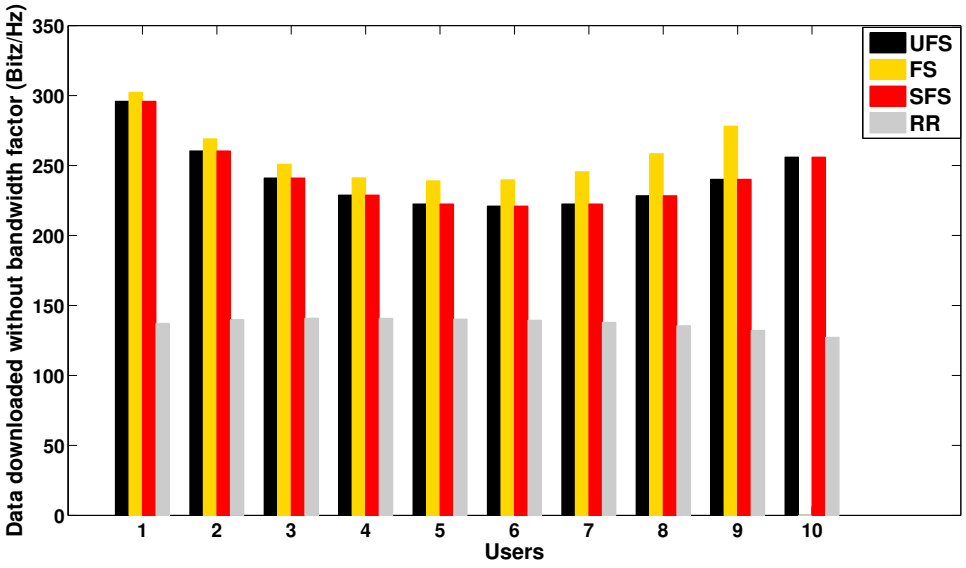


Figure 5.25: Impact of an arrival in the beginning of the FS duration for the 4 schedulers

have enough time as in the first case to re-allocate the same data quantity between users, especially for the arriving user (user 10). This difference is more apparent in the last case were the user arrives at pratically the end of the scheduler namely at time $t = 380\ it$ (see Fig.5.27).

For the last two arrival times ($t = 200\ it$ and $t = 380\ it$), another difference is depicted. UFS and the SFS are a little bit different from each other. One can notice the difference especially for users from 4 to 9 of Fig.5.27. This shift is due to the fact that the UFS has not had time to correct the fairness between users comparing to the SFS which expected this arrival at this particular time.

Note that in the three cases (Figs.5.25-5.27), the results for the basic FS (yellow bars) remain unchanged since only the 9 oldest users present in the network are considered during all the time interval T. The 10th user needs to wait till the next scheduling interval T to be served.

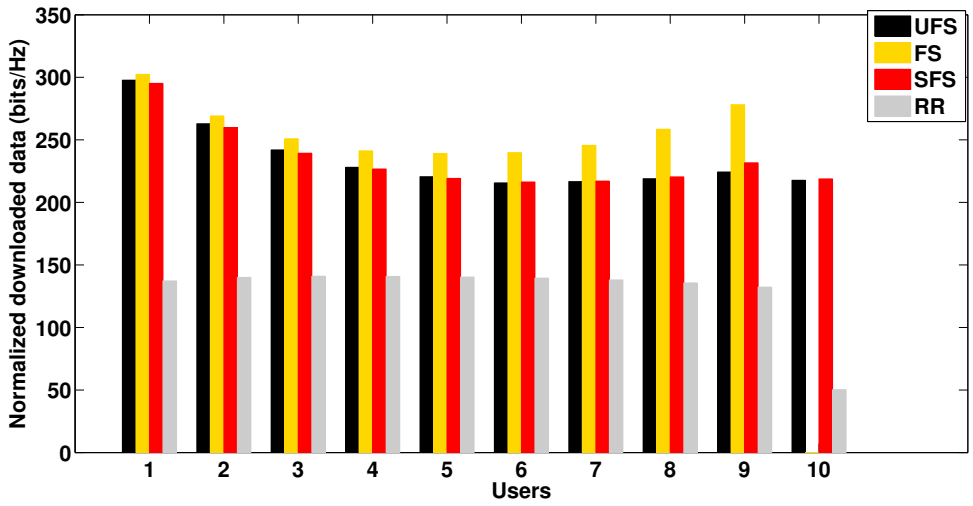


Figure 5.26: Impact of an arrival in the middle of the FS duration for the 4 schedulers

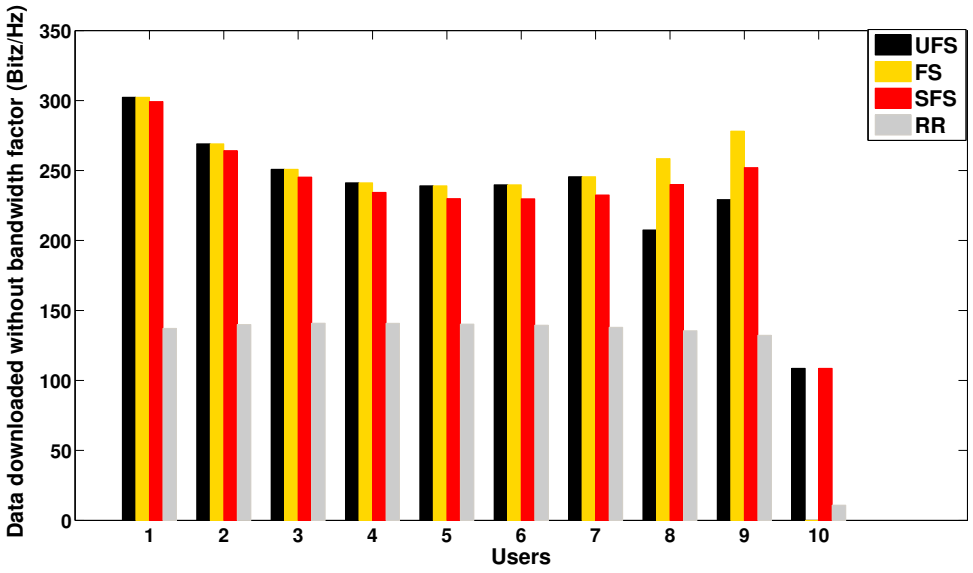


Figure 5.27: Impact of an arrival in the end of the FS duration for the 4 schedulers

Summary

We have introduced the forecast scheduling which is a novel scheduling

approach for users in high mobility that utilizes geo-localized measurements. Such measurements can be generated by a REM thanks to MDT data. This scheduling model consists of exploiting predicted SINR variations along the users' trajectories and allocate resources fairly between the users during the scheduling period. The model is based on maximizing a convex utility function under constraints and depends on a α fairness parameter. In this study, the CVX solver has been used to solve this optimization problem. We also proposed an analytical solution considering two users using the KKT. For the general cases, two heuristic algorithms are proposed to reduce the computing time of the forecast scheduling. The gain for both forecast scheduling and two best users heuristic algorithm are significant for all scenarios depicted in these thesis.

However a random event such as a variation in trajectories or a users' arrival/departure may occur during this scheduling horizon. The forecast scheduling is therefore updated to integrate past received data as soon as a random event occurs. This new scheduling is called the updated forecast scheduling (UFS).

Knowing the time and the type of the random event occurrence, the restricted forecast scheduling (RFS) is the best scheduler to use in so far as we implement the forecast scheduling only during a restricted time horizon.

We have shown firstly that the forecast scheduling model and the two best users algorithm can achieve very high MUT gains compared to RR scheduling. The gains remain significant also for moderate SINR variations along the trajectory. In the random cases, restricted and updated forecast schedulings outperform (depending on the random event type) comparing to the basic forecast scheduling and the RR scheduling. Higher and fair data downloaded for each user during their trajectories highlight this result.

Simulation scenarios have shown that errors in SINR measurements with respect to the predicted ones (e.g. from a REM) due to interference variations have negligible impact on the obtained throughputs. Such error may distort the forecast scheduling strategy (decisions). This work is just one example of how geo-localised measurements can improve RRM and optimize the network performance. Such measurements are likely to be available thanks to the MDT capable mobiles introduced in 4G networks.

Chapter 6

Conclusions and future works

6.1 Conclusion

In this thesis, we have proposed different ways to manage and control the service quality in the context of mobility in LTE Advanced and 5G networks. We have firstly introduced dynamic performance indicators for users in mobility. Secondly, we have studied the real life congestion and its mitigation thanks to different technologies: Small-cells deployment, VSC, and multilevel beamforming with global codebook with Large Scale Antenna System. Finally a resource management for users in high mobility is proposed.

To study the impact of the mobility in the network, we proposed different dynamic performance indicators: dynamic traffic and cell load, SINR outage probability, throughput outage probability, dynamic file transfer time and dynamic mean user throughput. These KPIs describe the evolution in time of the service quality in the network, and give an idea of service quality of the network. The main use case studied in this work to analyze the impact of mobility is the problem of the physical congestion. In general a physical congestion leads to radio congestion in the network. The particular case of a traffic light is studied. One can see, thanks to the dynamical KPIs introduced, that this congestion deteriorates the QoS of the users. Different type of technologies are proposed in this thesis to offload the macro cells and to better focus the transmitted signal on group of users. It has been shown that the different beam focusing techniques, namely virtual small cell (VSC), VSC combined with SON and the multilevel beamforming and its application to a heterogeneous antenna system (global codebook) can significantly improve quality of service. This trend will considerably evolve with the technological advances of massive MIMO in 5G networks in different frequency bands such as 3.5 GHz,

but also cm- and mm-wave bands, multi connectivity (e.g. combining LTE 800 MHz with submillimeter band), and others. At the longer term, energy consumption and cost of such technologies will reduce making it particularly attractive.

Finally, the possibility to benefit from long term opportunistic gain of users in high mobility has been investigated. This problem is of particular interest since with the increase of user speed, the coherence time diminishes, thus decreasing the opportunistic gain. The scheduling proposed in this thesis takes advantage of the future SINR variations along the users' trajectories, supposing that a REM is available and provides such metric. The scheduling is called the Forecast Scheduling. However, some random events can occur at any time such as an arrival, a departure, a change of trajectory. Two other forecast schedulings are proposed: Updated and Restricted Forecast Schedulings. These schedulings stop as soon as a random event may occur (next to the crossroad) or is occurring (arrival or departure of a user), and integrate the history of all users' data volume downloaded in the forecast scheduling computation. The main difference between UFS and RFS is that for the first scheduler we do not know the occurrence time of the event on the contrary of the RFS that knows the occurrence time. This contribution is among the first that shows the huge potential of using geo-localised measurements in improving and optimizing network performance and quality of service.

6.2 Future works

Users mobility, and in general their dynamicity, open a broad scope of 5G and next generations' works. Users dynamicity adds uncertainty to network performance indicators. This opens the field of machine learning and big data domain. The need to know better the past, present and future to better manage the resource between cells and users push several operators to create models and new technologies to follow each user's personal behaviour.

The work presented in this thesis opens the door to several challenging topics that are described presently.

Global multilevel codebook

- This work has introduced the concept of global multilevel codebook. This concept has been addressed from the modeling and performance point of view. The impact of this solution in terms of architecture of the antenna system is important and can be studied more deeply.

- The codebook construction for the use cases described in the thesis has been carried out manually. The automated construction of the multilevel codebook using basic geometric parameters of the targeted coverage area could be very useful for the large scale deployment of this solution. Automated codebook construction can be viewed as a self-configuration SON function. It could be performed at the base station or at a dedicated SON server at the management plane. Using machine learning methods, and particularly unsupervised learning including clustering, one can make the network learns which beam should be used directly for any user (fixed or mobile). This method will make a decision about the selection of the most appropriate beam between a cluster of beams.

Forecast Scheduling and advances antenna systems

5G technology will bring considerable advances in Large Scale Antenna System (LSAS), such as Massive MIMO in Time Domain Duplex (TDD) with non-codebook focusing capability or multi-connectivity. The combination of these technologies with the Forecast Scheduling could further increase the achieved performance and QoS gain. In the context of beam focusing forecasts, one is confronted with random variation of vehicles' speed. The focusing beam may not be centered on the vehicle's antenna. Two solutions can be considered: 1/ a mathematical solution that integrates random speeds on the FS and the predicted beam by using an extention of the FS using Markov Decision Process (MDP). 2/ A technological solution which could be more interesting and

optimal, is to integrate mobile antennas attached with a spring on the vehicles' roof. Antennas will therefore move with vehicles' kinetic depending on the spring's stiffness.

Forecast Scheduling applications with GLM

This thesis has presented one application of the use of a REM in deriving efficient scheduling algorithm. The use of a REM and in general, of geo-localized measurements in Radio Resource Management and SON algorithm is challenging and has a huge potential in improving performance and QoS of future networks. An example of such an application is the use of Forecast Scheduling in conjunction with SON functions that adapt network parameters, thresholds and resource allocation between groups of mobile users. One can use the Forecast Scheduling to optimize the resource sharing including the future SINR fluctuation for a cluster of mobile users. Arrivals, departures and any random event occurrence may be integrated by updating or restricting the resource sharing using the forecast scheduling.

In fact, all the antenna and LSAS technologies are already deployed in the networks. What remains is to deploy the appropriate, flexible and dynamic models in the networks to enhance not only the QoS but also the latency and the users' experience.

Chapter 7

Annexe

Proof of Theorem 1:

Suppose that the two users are scheduled together at two time instants K and h, then using the equivalence (5.10) and (5.13) thus:

$$\frac{\nu_K}{\phi(S_1^K)} = \frac{\nu_h}{\phi(S_1^h)}$$
$$\frac{\nu_K}{\phi(S_2^K)} = \frac{\nu_h}{\phi(S_2^h)},$$

therefore:

$$\frac{\nu_K}{\nu_h} = \frac{\phi(S_1^K)}{\phi(S_1^h)} = \frac{\phi(S_2^K)}{\phi(S_2^h)} \tag{7.1}$$

The data rates are almost surely not equal between users, as two users will never have two same positions in space at the same time, hence they cannot be scheduled simultanesouly in more than one time instance.

Proof of Theorem 3:

If we suppose that the two users (1 and 2) are scheduled at some time K i.e. $a_i(K) \neq 0$ with $i \in \{1, 2\}$, and that $T > 1$, then according to Theorem 2 for all times $t \neq K$ there is at most one user scheduled at time t. Let i_1 and i_2 be in $\{1, 2\}$, and for all $i \in \{1, 2\}$:

$$\begin{cases} (a_{i_1}(K)\phi(S_{i_1}^K) + \sum_{t=1,t\neq K}^{T} \phi(S_{i_1}^t)1_{\psi_{i_1}^t > \psi_{i_2}^t})^{-\alpha} = -\frac{\nu_K}{\phi(S_{i_1}^K)} \\ \sum_{i=1}^{2} a_i(K) = 1, \\ \lambda_{i,t} \geq 0, \\ \lambda_{i,t} a_i(t) = 0 \\ a_i(t) \geq 0. \end{cases}$$

Therefore:

$$a_{i_1}(K) = \frac{(-\frac{\nu_K}{\phi(S_{i_1}^K)})^{-\frac{1}{\alpha}} - \sum_{t=1}^{T} \phi(S_{i_1}^t)1_{\psi_{i_1}^t > \psi_{i_2}^t})}{\phi(S_{i_1}^K)} \tag{7.2}$$

We know that:

$$\sum_{i=1}^{2} a_i(K) = 1, \tag{7.3}$$

then we have:

$$\nu_K = -\frac{(\phi(S_1^K)^{1/\alpha-1} + \phi(S_2^K)^{1/\alpha-1})^{\alpha}}{(1 + \sum_{t\neq K} \psi_1^t 1_{\psi_1^t \geq \psi_2^t} + \psi_2^t 1_{\psi_2^t \geq \psi_1^t})^{\alpha}}. \tag{7.4}$$

We can now check if really a certain K exists or not by verifying whether $a_i(K) > 0$ for $i = 1, 2$ supposing the scheduling allocation for $t \neq K$ satisfies eq.5.20 in Theorem 2. If it is not the case then K does not exist. $a_1(K)$ and $a_2(K)$ are then expressed as follows:

$$a_1(K) = \frac{1}{\phi(S_1^K)}((\frac{-\nu_K}{\phi(S_1^K)})^{-1/\alpha} - \sum_{t\neq K} \phi(S_1^t)1_{\psi_1^t \geq \psi_2^t}) > 0 \tag{7.5}$$

$$a_2(K) = \frac{1}{\phi(S_2^K)}((\frac{-\nu_K}{\phi(S_2^K)})^{-1/\alpha} - \sum_{t\neq K} \phi(S_2^t)1_{\psi_2^t \geq \psi_1^t}) > 0, \tag{7.6}$$

with ν_K defined in equation (7.4). Therefeore $a_2(K) > 0$ is equivalent to:

$$\frac{1 + \sum_{t\neq K} \psi_1^t 1_{\psi_1^t \geq \psi_2^t} + \psi_2^t 1_{\psi_2^t \geq \psi_1^t}}{\phi(S_1^K)^{1/\alpha-1} + \phi(S_2^K)^{1/\alpha-1}} > \frac{\sum_{t\neq K} \phi(S_2^t)1_{\psi_2^t \geq \psi_1^t}}{\phi(S_2^K)^{1/\alpha}} \tag{7.7}$$

$$\iff$$

$$\frac{1 + \sum_{t\neq K} \psi_1^t 1_{\psi_1^t \geq \psi_2^t} + \psi_2^t 1_{\psi_2^t \geq \psi_1^t}}{\phi(S_1^K)^{1/\alpha-1} + \phi(S_2^K)^{1/\alpha-1}} > \frac{\sum_{t\neq K} \psi_2^t 1_{\psi_2^t \geq \psi_1^t}}{\phi(S_2^K)^{1/\alpha-1}} \tag{7.8}$$

We multiply the both sides of the inequality with the both denominators we have then:

$$\phi(S_2^K)^{1/\alpha-1} + \phi(S_2^K)^{1/\alpha-1} \sum_{t\neq K} \psi_1^t 1_{\psi_1^t \geq \psi_2^t}$$
$$-\phi(S_1^K)^{1/\alpha-1} \sum_{t\neq K} \psi_2^t 1_{\psi_2^t \geq \psi_1^t} > 0.$$

Using $\phi(S_i^K)\psi_1^t = \phi(S_i^t)$, and denote by ϕ_1 for $\phi(S_1^K)$ and ϕ_2 for $\phi(S_2^K)$ the last inequality is equivalent to:

$$-\phi_2^{\frac{1}{\alpha}} \sum_{t\neq K} \phi(S_1^t) 1_{\psi_1^t > \psi_2^t} + \phi_1^{\frac{1}{\alpha}} \sum_{t\neq K} \phi(S_2^t) 1_{\psi_2^t > \psi_1^t} < \phi_2^{\frac{1}{\alpha}} \phi_1$$

and $a_1(K) > 0$ is equivalent to:

$$\phi_2^{\frac{1}{\alpha}} \sum_{t\neq K} \phi(S_1^t) 1_{\psi_1^t > \psi_2^t} - \phi_1^{\frac{1}{\alpha}} \sum_{t\neq K} \phi(S_2^t) 1_{\psi_2^t > \psi_1^t} < \phi_1^{\frac{1}{\alpha}} \phi_2$$

If the inequalities (7.9) and (7) are verified then we have an expression for $a_1(K)$ and $a_2(K)$, and for all $t \neq K$ the user $i(t)$ scheduled is:

$$i(t) = argmax_{i\in\{1,2\}} \frac{\phi(S_i^t)}{\phi(S_i^K)}.$$

Proof of Theorem 5:

1/ We first demonstrate *"if user 1 is scheduled at time one then we have the inequatiy of Theorem 5":*

Using KKT condition with $\lambda_{2,1} \geq 0$ then for user 2:

$$\phi(S_2^1)(\sum_{j=1}^{T} a_2(j)\phi(S_2^j))^{-\alpha} + \lambda_{2,1} = -\nu_1, \tag{7.9}$$

and for user 1 $\lambda_{1,1} = 0$:

$$\phi(S_1^1)(\sum_{j=1}^{T} a_1(j)\phi(S_1^j))^{-\alpha} = -\nu_1, \tag{7.10}$$

if $\lambda_{2,1} > 0$, then

$$\frac{\phi(S_1^1)}{(\sum_{j=1}^{T} a_1(j)\phi(S_2^j))^\alpha} > \frac{\phi(S_2^1)}{(\sum_{j=1}^{T} a_2(j)\phi(S_2^j))^\alpha}, \tag{7.11}$$

or written differently:

$$\phi(S_1^1)(\sum_{j=1}^{T} a_2(j)\phi(S_2^j))^\alpha > \phi(S_2^1)(\sum_{j=1}^{T} a_1(j)\phi(S_1^j))^\alpha \tag{7.12}$$

Using the fact that if $\psi_1^t > \psi_2^t$ the user 1 is scheduled in time t. For sake of simplicity, we denote by $A_1 = A_1^1$ and $A_2 = A_2^1$, where $A_1^1 = \{t > 1, \psi_1^t > \psi_2^t\} \cup \{1\}$ and $A_2^1 = \{t > 1, \psi_1^t < \psi_2^t\}$. If $t \in A_1$ then user 1 is scheduled at time t, hence from inequality (7.12 we deduce:

$$\phi(S_1^1)(\sum_{A_2} a_2(j)\phi(S_2^j))^\alpha > \phi(S_2^1)(\sum_{A_2} a_1(j)\phi(S_1^j) + \sum_{A_1} \phi(S_1^j))^\alpha \tag{7.13}$$

with $a_2(j) = 1 - a_1(j)$.

Since $1 \geq a_2(j) \geq 0$, using these bounds:

$$\phi(S_1^1)(\sum_{A_2} \phi(S_2^j))^\alpha > \phi(S_2^1)(\sum_{A_1} \phi(S_1^j))^\alpha \tag{7.14}$$

This inequality must also be valid for all $k \in A_1$ if user 1 is selected at time $t = 1$:

$$\phi(S_1^k)(\sum_{A_2} \phi(S_2^j))^\alpha > \phi(S_2^k)(\sum_{A_1} \phi(S_1^j))^\alpha \tag{7.15}$$

Consider next the case where $\lambda_{2,1} = 0$ (rarely occurs when user 2 is not scheduled), therefore from equation (5.25):

$$\lambda_{1,t} - \lambda_{2,t} = \nu_1 \psi_1^{t/1} - \nu_1 \psi_2^{t/1}. \tag{7.16}$$

From the sign of (7.16), user 1 is scheduled if and only if $t \in A_1$ and user 2 is scheduled at all times $t \in A_2$. Using equations (7.9) and (7.10):

$$\frac{\phi(S_1^1)}{(\sum_{j \in A_1} \phi(S_1^j))^\alpha} = \frac{\phi(S_2^1)}{(\sum_{j \in A_2} \phi(S_2^j))^\alpha} \tag{7.17}$$

We can write:

$$\frac{\phi(S_1^t)}{(\sum_{j\in A_1}\phi(S_1^j))^\alpha} = \frac{\phi(S_1^1)}{(\sum_{j\in A_1}\phi(S_1^j))^\alpha} \times \frac{\phi(S_1^t)}{\phi(S_1^1)} \tag{7.18}$$

If $t \in A_1$ for $t \neq A_1$, we can write:

$$\frac{\phi(S_1^t)}{(\sum_{j\in A_1}\phi(S_1^j))^\alpha} > \frac{\phi(S_2^t)}{(\sum_{j\in A_2}\phi(S_2^j))^\alpha} \tag{7.19}$$

2/ We demonstrate the reciprocal *"if we have the inequality of the theorem 5 then the user 1 is scheduled at time 1"*:

$\Leftarrow$ we suppose that;

$$(\sum_{A_2}\phi(S_2^j))^\alpha > \frac{\phi(S_2^1)}{\phi(S_1^1)}(\sum_{A_1}\phi(S_1^j))^\alpha$$

and we suppose that the user 2 is scheduled at time 1, therefore:

$$\phi(S_1^1)(\sum_{A_2\cup\{1\}}\phi(S_2^j))^\alpha < \phi(S_2^1)(\sum_{A_1-\{1\}}\phi(S_1^j))^\alpha$$

But we have:

$$(\sum_{A_2\cup\{1\}}\phi(S_2^j))^\alpha > (\sum_{A_2}\phi(S_2^j))^\alpha$$

and:

$$\phi(S_2^1)(\sum_{A_1-\{1\}}\phi(S_1^j))^\alpha < \phi(S_2^1)(\sum_{A_1}\phi(S_1^j))^\alpha$$

Thus,

$$(\sum_{A_2}\phi(S_2^j))^\alpha < \frac{\phi(S_2^1)}{\phi(S_1^1)}(\sum_{A_1}\phi(S_1^j))^\alpha$$

Which is a contradiction, therefore the user 1 is scheduled if the inequality of theorem 5 is verified.

And the theorem is valid for all time u.

Bibliography

[1] A. Osseiran *et al.*, "Scenarios for 5G mobile and wireless communications: the vision of the METIS project," *Communications Magazine, IEEE*, vol. 52, no. 5, pp. 26–35, 2014.

[2] M. Fiore, J. Harri, F. Filali, and C. Bonnet, "Vehicular mobility simulation for vanets," in *Simulation Symposium, 2007. ANSS'07. 40th Annual*. IEEE, 2007, pp. 301–309.

[3] G. Araniti, C. Campolo, M. Condoluci, A. Iera, and A. Molinaro, "Lte for vehicular networking: a survey," *IEEE Communications Magazine*, vol. 51, no. 5, pp. 148–157, 2013.

[4] M. Gerla, C. Wu, G. Pau, and X. Zhu, "Content distribution in vanets," *Vehicular Communications*, vol. 1, no. 1, pp. 3–12, 2014.

[5] F. Morlot, S. Elayoubi, and F. Baccelli, "An interaction-based mobility model for dynamic hot spot analysis," in *INFOCOM, 2010 Proceedings IEEE*. IEEE, 2010, pp. 1–9.

[6] J. Broch, D. A. Maltz, D. B. Johnson, Y.-C. Hu, and J. Jetcheva, "A performance comparison of multi-hop wireless ad hoc network routing protocols," in *Proceedings of the 4th annual ACM/IEEE international conference on Mobile computing and networking*. ACM, 1998, pp. 85–97.

[7] B. Liang and Z. J. Haas, "Predictive distance-based mobility management for pcs networks," in *INFOCOM'99. Eighteenth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE*, vol. 3. IEEE, 1999, pp. 1377–1384.

[8] D. Helbing and B. Tilch, "Generalized force model of traffic dynamics," *Physical Review E*, vol. 58, no. 1, p. 133, 1998.

[9] D. C. Gazis, R. Herman, and R. W. Rothery, "Nonlinear follow-the-leader models of traffic flow," *Operations research*, vol. 9, no. 4, pp. 545–567, 1961.

[10] L. A. Pipes, "An operational analysis of traffic dynamics," *Journal of applied physics*, vol. 24, no. 3, pp. 274–281, 1953.

[11] W. Leutzbach, *Introduction to the theory of traffic flow*. Springer, 1988, vol. 47.

[12] M. Jepsen, "On the speed-flow relationships in road traffic: a model of driver behaviour," in *Third International Symposium on Highway Capacity*, no. Volume 1, 1998.

[13] I. Prigogine and R. Herman, "Kinetic theory of vehicular traffic," Tech. Rep., 1971.

[14] C. H. Foh, G. Liu, B. S. Lee, B.-C. Seet, K.-J. Wong, and C. P. Fu, "Network connectivity of one-dimensional manets with random waypoint movement," *IEEE Communications Letters*, vol. 9, no. 1, pp. 31–33, 2005.

[15] E. Atsan and Ö. Özkasap, "A classification and performance comparison of mobility models for ad hoc networks," in *Ad-Hoc, Mobile, and Wireless Networks*. Springer, 2006, pp. 444–457.

[16] D. R. Choffnes and F. E. Bustamante, "An integrated mobility and traffic model for vehicular wireless networks," in *Proceedings of the 2nd ACM international workshop on Vehicular ad hoc networks*. ACM, 2005, pp. 69–78.

[17] T. Bonald and A. Proutiere, "A queueing analysis of data networks," in *Queueing Networks*. Springer, 2011, pp. 729–765.

[18] M. K. Karray, "User's mobility effect on the performance of wireless cellular networks serving elastic traffic," *Wireless Networks*, vol. 17, no. 1, pp. 247–262, 2011.

[19] L. Fenton, "The sum of log-normal probability distributions in scatter transmission systems," *Communications Systems, IRE Transactions on*, vol. 8, no. 1, pp. 57–67, 1960.

[20] T. Bonald and A. Proutière, "Wireless downlink data channels: User performance and cell dimensioning," in *Proc. of ACM Mobicom*, 2003.

[21] 3GPP, "Evolved Universal Terrestrial Radio Access (E-UTRA); Mobility enhancements in heterogeneous networks (Release 11)," 3rd Generation Partnership Project (3GPP), TR 36.839, Dec. 2012.

[22] A. Tall, Z. Altman, and E. Altman, "Self organizing strategies for enhanced icic (eicic)," *Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks (WiOpt), 2014 Proceedings of the 12th International Symposium on*, 2014.

[23] J. Mo and J. Walrand, "Fair end-to-end window-based congestion control," *IEEE/ACM Transactions on Networking (ToN)*, vol. 8, no. 5, pp. 556–567, 2000.

[24] M. Benaïm and N. El Karoui, *Promenade aléatoire: Chaînes de Markov et simulations; martingales et stratégies*. Editions Ecole Polytechnique, 2005.

[25] S. Hur *et al.*, "Multilevel millimeter wave beamforming for wireless backhaul," in *Proc. IEEE GLOBECOM Workshops (GC Wkshps), 2011*. IEEE, 2011, pp. 253–257.

[26] A. Osseiran *et al.*, "Scenarios for 5G mobile and wireless communications: the vision of the METIS project," *IEEE Communications Magazine*, vol. 52, no. 5, pp. 26–35, 2014.

[27] A. Galindo-Serrano, S. Martinez Lopez, and A. Gati, "Virtual small cells using large antenna arrays as an alternative to classical HetNets," in *First International Workshop on Intelligent Design and Performance Evaluation of LTE-Advanced Networks*, Glasgow, Scotland, May 2015.

[28] A. Tall, Z. Altman, and E. Altman, "Virtual sectorization: design and self-optimization," in *5th International Workshop on Self-Organizing Networks (IWSON 2015)*, Glasgow, Scotland, May 2015.

[29] ——, "Multilevel beamforming for high data rate communication in 5G networks," *arXiv preprint arXiv:1504.00280*, 2015.

[30] 3GPP, "Evolved Universal Terrestrial Radio Access (E-UTRA); Radio Frequency (RF) system scenarios," 3rd Generation Partnership Project (3GPP), TR 36.942, Sep. 2012.

[31] P. Viswanath, D. N. Tse, and R. Laroia, "Opportunistic beamforming using dumb antennas," *IEEE Transactions on Information Theory*, vol. 48, no. 6, pp. 1277–1294, 2002.

[32] U. Charash, "Reception through nakagami fading multipath channels with random delays," *IEEE Transactions on Communications*, vol. 27, no. 4, pp. 657–670, 1979.

[33] W. Bao and S. Valentin, "Bitrate adaptation for mobile video streaming based on buffer and channel state," in *2015 IEEE International Conference on Communications (ICC)*. IEEE, 2015, pp. 3076–3081.

[34] N. Bui, S. Valentin, and J. Widmer, "Anticipatory quality-resource allocation for multi-user mobile video streaming," in *2015 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*. IEEE, 2015, pp. 245–250.

[35] I. Triki, R. El-Azouzi, and M. Haddad, "Anticipating resource management and qoe provisioning for mobile video streaming," *arXiv preprint arXiv:1512.05705*, 2015.

[36] F. Capozzi, G. Piro, L. A. Grieco, G. Boggia, and P. Camarda, "Downlink packet scheduling in lte cellular networks: Key design issues and a survey," *IEEE Communications Surveys & Tutorials*, vol. 15, no. 2, pp. 678–700, 2013.

[37] R. Jain, "Channel models a tutorial1," 2007.

[38] R. B. Ertel, P. Cardieri, K. W. Sowerby, T. S. Rappaport, and J. H. Reed, "Overview of spatial channel models for antenna array communication systems," *IEEE Personal Communications*, vol. 5, no. 1, pp. 10–22, 1998.

[39] J. Yao, S. S. Kanhere, and M. Hassan, "Improving qos in high-speed mobility using bandwidth maps," *IEEE Transactions on Mobile Computing*, vol. 11, no. 4, pp. 603–617, 2012.

[40] W. Hapsari, A. Umesh, M. Iwamura, M. Tomala, B. Gyula, B. Sebire *et al.*, "Minimization of drive tests solution in 3GPP," *IEEE Communications Magazine*, vol. 50, no. 6, pp. 28–36, 2012.

[41] H. Braham, S. Ben Jemaa, B. Sayrac, G. Fort, and E. Moulines, "Coverage mapping using spatial interpolation with field measurements," in *2014 IEEE 25th Annual International Symposium on Personal, Indoor, and Mobile Radio Communication (PIMRC)*. IEEE, 2014, pp. 1743–1747.

[42] A. Galindo-Serrano *et al.*, "Cellular coverage optimization: A radio environment map for minimization of drive tests," in *Cognitive Communication and Cooperative HetNet Coexistence*. Springer, 2014, pp. 211–236.

[43] N. Cressie and G. Johannesson, "Fixed rank kriging for very large spatial data sets," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 70, no. 1, pp. 209–226, 2008.

[44] E. Papkelis, I. Psarros, L. C. Ouranos, C. G. Moschovitis, K. Karakatselos, E. Vagenas, H. Anastassiu, and P. Frangos, "A radio-coverage prediction model in wireless communication systems based on physical optics and the physical theory of diffraction [wireless corner]," *IEEE Antennas and Propagation Magazine*, vol. 49, no. 2, pp. 156–165, 2007.

[45] Q. Gong, Y. Li, and Z.-R. Peng, "Trip-based optimal power management of plug-in hybrid electric vehicles," *IEEE Transactions on Vehicular Technology*, vol. 57, no. 6, pp. 3393–3401, 2008.

[46] B. M. Williams and L. A. Hoel, "Modeling and forecasting vehicular traffic flow as a seasonal arima process: Theoretical basis and empirical results," *Journal of transportation engineering*, vol. 129, no. 6, pp. 664–672, 2003.

[47] W. H. Lam, Y. Tang, K. Chan, and M.-L. Tam, "Short-term hourly traffic forecasts using hong kong annual traffic census," *Transportation*, vol. 33, no. 3, pp. 291–310, 2006.

[48] J. You and T. J. Kim, "Empirical analysis of a travel-time forecasting model," *Geographical Analysis*, vol. 39, no. 4, pp. 397–417, 2007.

[49] Y.-J. Yu and M.-G. Cho, "The system for predicting the traffic flow with the real-time traffic information," *Journal of the Korea Institute of Information and Communication Engineering*, vol. 10, no. 7, pp. 1312–1318, 2006.

[50] I. Steven, J. Chien, and C. M. Kuchipudi, "Dynamic travel time prediction with real-time and historic data," *Journal of transportation engineering*, 2003.

[51] T. Liu, P. Bahl, and I. Chlamtac, "Mobility modeling, location tracking, and trajectory prediction in wireless atm networks," *IEEE Journal on Selected Areas in Communications*, vol. 16, no. 6, pp. 922–936, 1998.

[52] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.

[53] P. Selvidge, "How long is too long to wait for a website to load," *Usability news*, vol. 1, no. 2, 1999.

[54] ——, "Examining tolerance for online delays," *Usability News*, vol. 5, no. 1, pp. 1–5, 2003.

[55] R. B. Miller, "Response time in man-computer conversational transactions," in *Proceedings of the December 9-11, 1968, fall joint computer conference, part I.* ACM, 1968, pp. 267–277.

[56] J. Nielsen, "Response times: the three important limits," 1994.

[57] M. Pavone, N. Bisnik, E. Frazzoli, and V. Isler, "A stochastic and dynamic vehicle routing problem with time windows and customer impatience," *Mobile Networks and Applications*, vol. 14, no. 3, p. 350, 2009.

[58] D. Hendricks, "Evaluation of value-at-risk models using historical data (digest summary)," *Economic Policy Review Federal Reserve Bank of New York*, vol. 2, no. 1, pp. 39–67, 1996.

[59] T. J. Linsmeier and N. D. Pearson, "Value at risk," *Financial Analysts Journal*, vol. 56, no. 2, pp. 47–67, 2000.

[60] H. P. Palaro and L. K. Hotta, "Using conditional copula to estimate value at risk," *Journal of Data Science*, vol. 4, pp. 93–115, 2006.

[61] D. Fantazzini, "Dynamic copula modelling for value at risk," *Frontiers in Finance and Economics*, vol. 5, no. 2, pp. 72–108, 2008.

[62] S. Sethi and G. Sorger, "A theory of rolling horizon decision making," *Annals of Operations Research*, vol. 29, no. 1, pp. 387–415, 1991.

[63] I. CVX Research, "CVX: Matlab software for disciplined convex programming, version 2.0," http://cvxr.com/cvx, Aug. 2012.

[64] M. Grant and S. Boyd, "Graph implementations for nonsmooth convex programs," in *Recent Advances in Learning and Control*, ser. Lecture Notes in Control and Information Sciences, V. Blondel, S. Boyd, and H. Kimura, Eds. Springer-Verlag Limited, 2008, pp. 95–110.